University of Alberta

The Rationality of Self-Revolutions

by

Paul Michael Barrette

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Arts

Department of Philosophy

Edmonton, Alberta
Fall 2008

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-46948-4
Our file Notre référence
ISBN: 978-0-494-46948-4

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

An analysis of Robert Nozick's closest continuer theory of personal identity (1981) shows that, without some further assumption, the closest continuer theory of personal identity cannot plausibly claim that radical changes in persons preserve personhood. Nozick's theory of value (1981) enables a distinction between valuable and non-valuable radical changes in persons. Nozick's notion of 'symbolic utility' (1993) shows how a valuable change can be rational. With the rationality of valuable changes, the determinacy of personal identity over radical changes survives Derek Parfit's (1984) objection that personal identity is only a matter of degree. I conclude that personal identity is not equally a matter of degree in all cases, but that it can become a more determinate entity, the continuation of which is less a matter of degree. Personal identity becomes more determinate under value-increasing self-syntheses and less determinate to the extent that one's self-syntheses are value-decreasing.

ACKNOWLEDGEMENTS

I Thank Wes Cooper for his inspiring class on personal identity
and Adam Morton for his editing and his many suggestions.

DEDICATION

Dedicated to Kayla, who gave me ideas and companionship throughout, and to my grandparents, on whose bookshelf I found a pathway into philosophy through Plato's Dialogues.

TABLE OF CONTENTS

I. The Closest Continuer Theory

II. Self-Revolutions

III. Not All Self-Revolutions Preserve Personhood

IV. Value Theory

V. Decision and Value

VI. Degrees of Determinacy

VII. The Rationality of Self-Revolutions

VIII. Reductionism and the Determinacy of Identity

IX. Rationality and Reductionism

X. Bibliography

In *Philosophical Explanations* (1981), Robert Nozick develops both his closest continuer theory of personal identity and his objective theory of value as organic unity. I show that, without some further assumption, the closest continuer theory of personal identity cannot plausibly claim that radical changes in persons preserve self-identity. Using Nozick's theory of value, I distinguish valuable from non-valuable radical changes in persons and show that, to the extent that a change is valuable, it preserves and increases one's personhood. Using Nozick's notion of 'symbolic utility' (1993), I show that a valuable change can be a rational change. Under added value-preserving or value-increasing conditions, the determinacy of personal identity over radical changes is rid of implausible consequences and can survive Derek Parfit's (1984) reductionist objection that personal identity is only a matter of degree. I argue that personal identity is not equally a matter of degree in all cases, but that it can become a more determinate entity, the continuation of which is less a matter of degree. As I explain below, personal identity becomes more determinate under self-syntheses that track value and becomes less determinate to the extent that one's self-syntheses do not track value.

I. The Closest Continuer Theory

Nozick construes the self-identity of a person in terms of a more general theory of identity between objects. This more general theory uses the 'closest relation schema,' which he derives from his notion of an informative classification. When categorizing objects, an informative classification is one that minimizes the difference between members of the same subclasses and maximizes the difference between members of different subclasses (1981 p. 84). An informative classification brings about what he calls an 'organic unity'. An organic unity occurs when as much diversity as possible is brought

into a significant unity (p. 86). The meaning of 'significant' will depend upon the purpose of the particular classification. This explains why, as Nozick points out, a classification's informativeness is not preserved over isomorphism (p. 427). To demonstrate this, he uses the example of a phonograph record, which is isomorphic to the musical performance it records (p. 425). Since the degree of importance of the unifying relationships among the sounds and tones of the performance does not transfer to the unifying relationships among the microscopic spatial contours of the record, the significance of the unifying relations is not preserved over the isomorphism (p. 425).

One reaches an informative classification if it is performed, for some metric along which to measure closeness, while adhering to what Nozick calls the 'closest relation schema.' The schema ensures that "two things are part of the same class when they are close enough and there is no third thing not in the class which is closer to one of them than each and every other thing in the class is" (p. 85). Transitive identity uses the closest relation schema to build entities for a single time; it is the maximally informative classification in a single time-slice of space (p. 85). Longitudinal identity uses it to classify stages of the same entity to unify an object over time (p. 85). Together, transverse and longitudinal identities constitute what Nozick calls 'entification', which is the creation or delineation of entities (p. 85).

To establish a criterion for delineating the entity that a person's self-identity isolates, Nozick first imagines a person without self-identity. He does this by imagining a case of acts without a doer and then addresses the question of how self-identity can arise out of such acts (p. 88). One way it could arise through one of these acts is if dimensions and components of that act make certain dimensions and features salient for the grouping

of things as similar or relevantly close to itself (p. 88). This can occur in two steps. First, a number of acts classify, demarcate entities, apply the closest relation schema, etc (p. 88). Second, another 'reflexively self-referring' act unifies all of these acts and itself into a single and unified whole (p. 89).[1] Linguistic devices provide one type of reflexively self-referring act. They are indexical terms with references that vary with the context of their utterance and depend essentially upon the utterance in which they appear (p. 74). For example, a person reflexively self-refers when she utters, 'me', or, 'this very person who is writing this sentence'. Using the dimensions and components of itself as the similarity standards for an application of the closest relation schema, the act classifies itself as part of the same entity as those features it marks as similar, and this entity becomes the self (p. 90). The self becomes the tightest organic unity including the act and, because the act is the very source of its unity, the act reflexively refers to it. This 'self-synthesis' occurs not only transversely, but longitudinally so as to include past entities, including former self-syntheses (p. 91).

The dimensions determining measures of closeness and sameness in a self-synthesis are those aspects of the world that the synthesized personhood cares about. Even if fundamental, they are not necessarily those that the pre-synthesized personhood (the personhood of a previous synthesis) cared about. This change in fundamental cares is possible because the act of self-synthesis is a self-subsuming decision,[2] a decision that

---

[1] Nozick acknowledges that there will be ways of unifying these acts without attributing them to a doer, but bids the reader to bear with him (p. 89). *This* way of unifying the acts, he will claim, can explain how reflexive self-awareness is possible and why the self is essentially a self (p. 91).

[2] Nozick discusses self-subsuming decisions at length in his chapter, "Choice and Indeterminism", in which he espouses a version of libertarianism; however, this model of decision-making does not depend upon indeterminism. Weight-bestowal need not depend upon having been able to bestow otherwise, it needs just, 'had I been able to bestow otherwise, this bestowal also would have been justified'. In his section on determinism, he mentions the possibility of a deterministic model (pp. 352-353).

motivates and justifies itself. Nozick arrives at the notion of a self-subsuming decision through noting that, whichever act one performs, different background conditions to that act exist, each of which can be raised to causal status. For example, if I choose A it will be because of the reasons for doing A, if I choose B it will be because of the reasons for doing B (pp. 294-295). He claims that, where a reason's weight is its importance in determining future actions or thoughts, decisions establish inequalities in weight among the reasons for different actions (even if the weights are not precise) (p. 294-297). A person does not discover weights but assigns them. Though assigned, the weight bestowal is not arbitrary because it is self-subsuming, that is, it fixes general principles mandating not only the act of its own execution but also the bestowal of weights (and similar bestowals of weights) that justify its execution (p. 300). Bestowed weights set up a framework for future decisions in a way that resembles a legal system where "the decision represents a tentative commitment to make future decisions in accordance with the weights it bestows" (p. 297). Thus, weight bestowal partially explains the dynamics behind both self-syntheses on the precedent of earlier self-synthesis and those that occur without precedent. I call those that occur without precedent 'revolutionary self-syntheses' (fully defined below).

Nozick finds an example of the self-subsuming decision in Herbert Simon's satisficing model of decision. It holds that an agent will do an action that is 'good enough', but if no action among the alternatives available meets the criterion, she will search for others; to the extent that her searches continue to fail, her criterion for what is 'good enough' becomes less ambitious (p. 300). Nozick embeds these considerations into an optimizing model of the cost of searching for new alternatives, the cost of gathering

information, and estimates about the probability of finding a new better alternative, where searching for new or more information about alternatives is always a readily available option (p. 300). He asks if choosing whether to make a decision or to keep searching depends upon a previous optimizing decision or upon a satisficing decision that the structuring was 'good enough'. He answers that, whichever of the two, the decision "includes estimates of the costs and benefits of gathering more information in *that very* choice situation", and that its scope covers all costs including its own; thus, it is a self-subsuming decision (p. 300).

Not every reflexive act of self-synthesis will create an entirely new set of past selves nor will a present synthesis determine the precise character of a future synthesis (p. 89). There is a spectrum extending from self-identifying on precedent to revolutionary self-identification. At one extreme, successive acts of self-synthesis on precedent will group things as similar along unchanging dimensions of sameness, that is, along those same dimensions by which one has constituted oneself the time before. At the other, I call a self-synthesis 'revolutionary' if it is unprecedented, that is, if it judges that one is still oneself along completely new dimensions of sameness than those used before. If one sets a metric for determining closeness by assigning weights to dimensions of closeness, then self-synthesis on precedent inherits the commitment of previous weight-bestowals and a revolutionary self-synthesis re-assigns weights.

For Nozick, a 'continuer' of X grows out of X's properties, is causally produced by them, or is explained by X's earlier having had its properties (p. 35). Because the dimensions of closeness that determine a person's metric are those assigned the most weight, Nozick points out that a continuer's closeness to X is often proportional to the

degree of care X has for that continuer (p. 63).[3] This means that to the extent that someone *could be* me (given my metric) they are called a 'continuer'. This is because that potential person would be, under some distribution of properties (potential resources for a self-synthesis), available for me to synthesize. For example, if I *would* synthesize as a healthy person and thereby identify as healthy, but my social situation requires that I work in unsafe environmental conditions and for such little wages that I cannot afford nutritional food, then I cannot synthesize as this. But if my social situation was less bleak and the appropriate resources were available to me, then I could identify as healthy. So the potentially healthy me is, speaking from a modal perspective, a continuer because of my property of having the metric I do, one that values healthiness. On the *closest* continuer theory, the closest of my potential continuers to my current values, given my metric, *will be* me.

It is probable that, in every day life, most syntheses occur on precedent with only minor changes to those dimensions of sameness by which one goes on being oneself. On the other hand, commonly held views, history, and myths reveal examples far to the revolutionary side of the self-identity spectrum. For example, Buddha's departure from his affluent lifestyle to seek enlightenment, Malcolm X's elevation from petty-crime to a leader in the fight for civil liberties, Saul's revelation on the road to Demascus, the 'born again' Christian, and the liberal precept that high-offending criminals can be rehabilitated. The people in these examples radically change, but intuition tells us they remain the same person as they were before their change. Prince Gautama's privileged upbringing is an essential aspect to the significance of his spiritual life wherein he

[3] Cases where they may not be proportional are if the care is for some stage of a relative of X or someone on X's side in a conflict or sharing some of X's properties (p. 661, f. 25).

renounces the importance of the ego. It is unimpressive if a born pauper rejects the material world; he doesn't even know what he's rejecting. Malcolm Little's initial petty-criminal activity, a product, he admits, of his socially stratified environment (*Autobiography* 1964, p. 384), is an essential aspect of his later revolutionary movement against this structure:

> "I believe that it would be almost impossible to find anywhere in America a black man who has lived further down in the mud of human society than I have; or a black man who has been any more ignorant than I have been; or a black man who has suffered more anguish during his life than I have. But it is only after the deepest darkness that the greatest light can come; it is only after extreme grief that the greatest joy can come; it is only after slavery and prison that the sweetest appreciation of freedom can come" (p. 385)

Likewise, Saul's history of trying to wipe out Christianity is essential to his narrative as the apostle of Christ, who gives the gift of grace and the forgiveness of all sins:

> For ye have heard of my conversation in time past in the Jews' religion, how that beyond measure I persecuted the church of God, and wasted it: And profited in the Jews' religion above many my equals in mine own nation, being more exceedingly zealous of the traditions of my fathers. But when it pleased God, who separated me from my mother's womb, and called me by his grace, To reveal his Son in me, that I might preach him among the heathen; immediately I conferred not with flesh and blood: Neither went I up to Jerusalem to them which were apostles before me; but I went into Arabia, and returned again unto Damascus... Afterwards I came into the regions of Syria and Cilicia; And was unknown by face unto the churches of Judaea which were in Christ: But they had heard only, That he which persecuted us in times past now preacheth the faith which once he destroyed. And they glorified God in me (Galatians 1:13-24).

The significance of these people's pre-revolutionary stage consists in what their post-revolutionary stage means to both others and, as we can see from these personal testimonies, to themselves. In this respect, their later stage *continues* their earlier stage

because the earlier stage plays a crucial role in their reasons for coming to care for and continuing to choose their self-identity as it is in their later stages.

Nozick's closest continuer theory of self-identity allows for revolutionary changes such as these to occur but it also allows for revolutionary changes that are unlike these, regressive and nonsense self-identifications that, as I explain below, do not rightfully gain worth or significance from pre-revolutionary identifications. In the next sections, I argue this point and give a more detailed analysis of revolutionary self-syntheses.

II. Self-Revolutions

Over time, Nozick claims, self-syntheses along the same dimensions and features "will happen repeatedly *in the absence of reason to deviate and override the precedents*" (1981 p. 89).[4] On precedence, self-synthesis merely maintains one's constitution and does not add anything to it. The only sort of decision that can override fundamental dimensions holding precedent is a self-subsuming decision. In a self-subsuming decision, one does not act on the pre-existing strongest preference, but a preference becomes strongest in the process of making a fundamental decision based on itself (p. 298). For example, imagine a Jewish person in the holocaust knows that if he can convince the Nazis that he is not Jewish, but himself a Nazi, this will enable the validation his family's passports. However, he was born with the inability to lie convincingly. The only way for him to be successful is to *become* a Nazi. In the act of performing his decision to become a Nazi, his fundamental cares and attitudes change and so does his metric of closeness. He no longer identifies with things like religion and family, but groups the Fuhrer and a

[4] My emphasis

hate for Hebrews in the act of synthesis and this makes him into a 'true blue' Nazi.[5] In this case, just what continuity *is* becomes a question; the cares that direct his attention, his thoughts, and his actions toward the aspects of life with which he identifies become usurped by new fundamental attitudes that will lead his life to trace new aspects. Thus, there is a break in continuity between sets of dimensions for determining sameness when applying the closest relation schema.

Nozick's stance on whether the Jewish person's self-revolution preserves self-identity is ambiguous, but, as I point out below, his view seems to imply that self-identity is always preserved over any revolution. I will argue that there are some types of revolution where self-identity is preserved and some types where it is not. First, however, I consider Nozick's position more closely. When a person's fundamental attitudes begin to break down as a metric for measuring the closer of one or more potential continuers, Nozick proposes one view the world through what he calls 'Platonic glasses.'

Nozick introduces the concept of Platonic glasses in response to the 'branching problem' with using closeness as a criterion of personal identity. If two different continuers can tie for closeness, e.g. if Y at t2 and Z at t2 both continue X at t1 with equal closeness, then this violates the transitivity of identity, that is, there can be no sense in which one person can be the same person as *both* of these two distinct persons (p. 62). Imagine walking into a transporter from Star Trek, it copies the structure of your body, disintegrates you, and then reproduces an exact copy of you. Except, this transporter malfunctions because it produces two copies of you – this is a tie. Moreover, problems with transitivity of identity aside, ties between continuers raise problems for concepts like

[5] A change like this occurs in the film, "The Believer" (2001).

'life' and 'death.' If one person is identical with two continuers and one continuer dies, it seems like the person both dies and continues to live at the same time (p. 62).

Nozick's solution to these problems stems from two ways that caring about something might function. One way is the Platonic mode, "you care about C [a concept] and this care spills over (or you transfer it) to C's best instantiated relation" (p. 67). The best instantiated relation obtains when no other instantiated relation R' is an as good or better realizer of concept C than R is (pp. 53-54). For example, if C is a person's self-identity, then the condition for its application is that X (a potential continuer) stand to *a* (a person) in the best instantiated relation e.g. you step into a transporter, it produces two copies, one is normal but one of them is inside-out, then the normal copy is you. Through Platonic glasses, the dimensions of sameness and difference in an act of self-synthesis delineate realms of caring by specifying modally the sorts of things one *would* care to identify with and deciding, given the actual context, what option best embodies these sorts of things e.g. in any context I would care for a healthy body. One measures the closeness of a continuer in accordance with where it stands on these preset dimensions (p. 69). The point is that one first decides what one cares about and then looks to the world to see, given what one has decided, how much one cares about the things in it. To arrive at the Platonic glasses mode of caring, one looks at the things in the world and imagines how they could be better. For example, a woman eats an apple for the first time and it has a bruise. The woman now has the concept 'apple', to which she assigns a degree of care. On the Platonic Glasses Mode of caring, by mental processes that involve realizing the apple is a fruit and deducing that all fruits have a non-bruised form, she imagines a fresh non-bruised apple and assigns a degree of care to *it*. Next she looks for the best

instantiated realization of her imagined concept which is, in this case, the bruised apple. Her care for the bruised apple then gets a relatively lower degree of care than the non-bruised imaginary apple.

The Best Realization mode, on the other hand, looks first at the best instantiated realization of a concept in order to see how much to care about *it* and only then decides how much any concept that this realization might fall under is worth caring about (p. 67). It is called the 'best realization' mode because only objects, never the concepts they might be instances of, determine assignments of value; under any construal of what constitutes the best realization of a concept, the care assigned to it never exceeds that assigned to its best realization. By this mode, the care the woman assigns to her concept, 'apple', does not exceed a due proportion to her enjoyment of the very apple that she ate. The best realization mode decides how much to care about a thing apart from whatever concept it might realize, and then the worth assigned to any concept it *does* realize depends upon the worth assigned this concept's best realization.

One can switch between the two modes of caring depending upon what is in focus. In the Best Realization mode, if there is no closest continuer but only a tie, then one merely assigns an amount of care to this tie and does not care that there is no one closest continuer (pp. 67-68). In contrast, if one cares about one's mono-continuity, the Platonic Glasses mode does care to some extent that that there is a closest continuer because this *would* realize identity better than the tie does (p. 68). Whether one does care about mono-continuity can vary from person to person. If a person adopts the Best Realization mode of caring, then he cares only about what, among the potential realizers readily available, could realize identity. If nothing realizes identity, one no longer cares

about identity, e.g. an atheist who is comfortable at the thought of her own death, in this case, utilizes the Best Realization mode of caring. On the other hand, the belief in an afterlife allows one to maintain the Platonic Glasses mode of Caring about identity at death through conceiving of and caring for ways that identity *could* survive bodily death. Whether one assumes the Platonic Glasses or Best Realization mode of caring can vary from person to person and, within a person, from time to time and from topic to topic (p. 67). A person fixes his degree of closeness to potential identifications in accordance with how much he cares about those potentialities (p. 69). Even within a mode of caring, the content of one's measure of closeness can vary from person to person and, within a person, from time to time (p. 69).

Consider the implications of applying either of these modes of caring in a self-synthesis. An application of the Platonic Glasses mode of caring involves considering the things one *would* care most about if they *were* available for self-synthesis, e.g. for me, these are healthiness, intelligence, courageousness, friendship etc. These things one would care about constitute one's metric. Of the things that *are* available for classification under the self-synthesis when it occurs, those that score high enough on the metric are included in the synthesis, e.g. a body, regular exercise, studying, a willingness to engage scary situations to eliminate fear, good friends, etc. Now imagine that the friends available to me are only somewhat good so that they are generally nice people and are nice to me, but they will never do the things that I want to do when our interests conflict. Since no other friends are available, the friends will do under the Best Realization mode of caring; however, if I am utilizing the Platonic glasses mode of

caring, I perform a self-synthesis where I identify with an idea of what a friend ought to be and judge prospective friends by their conformity to the idea.

An application of the Best Realization mode of caring involves *deriving* a metric from the sorts of things available for classification and not from what *would* care about were things different e.g. I identify with the friends I have and since I don't bother imagining better friends I would not be 'on the lookout' to identify with different friends; should I, by circumstance, make better friends, they would take precedence, but my care for friendship does not exceed my best memories and best experiences of friendship. I measure the value of a given friendship, in this mode, as relative to the one's I've actually experienced, not those that I can imagine. This is not to say that either mode of caring is better for friendship, each mode of caring has virtues and drawbacks, e.g. the first involves integrity but a lack of loyalty, the second involves loyalty but a lack of integrity. The point is that they are different ways of setting one's metric in a self-synthesis.

I claim that only a self-synthesis utilizing the Platonic Glasses mode of caring can preserve self-identity over a self-revolution. This is partially because the Platonic Glasses mode of caring allows for the possibility of 'progressive' self-syntheses. For example, the optimal realization of a person may require a complete break with that person's previous fundamental attitudes. An alcoholic whose fundamental care is to feel ethanol in his brain will require a break with this deep attitude to become sober. Setting aside for now the issue of why sobriety is better than alcoholism, to assign a degree of care to sobriety that overrides the care for tending an addiction, the alcoholic cannot simply *observe* that sobriety is better as could the woman observe her apple. The *only* way he can come to care for sobriety more is by *imagining* what it is like and then, in a revolution of his

preferences, choose the content of this imagination as more fundamentally valuable than drunkenness. It is possible that he could *remember* the sobriety he had before he became an addict, but then the representation of sobriety would have already existed in his memory and thus would have already fed into his current metric of closeness as determined by his current fundamental attitudes. Perhaps he could recall sobriety 'as if remembering it for the first time'. However, its lack of continuity with the original sobriety would render it little different from an imagination. If revolutionary self-syntheses are to be possible without ending one person and beginning another, in which case certain important sorts of 'personal progress' could not occur, they must utilize the Platonic Glasses mode of caring.

A self-revolution that ends one person and begins another occurs under the Best Realization mode of caring. If the alcoholic does not initiate and choose his own sobriety, then a fundamental change in his attitudes so that he values sobriety higher than drunkenness will constitute a break in his self-identity. For example, if he is hypnotized into changing his attitudes, then one cannot say that his identity is preserved over the change because the change did not stem from his prior identity but undermined it. Take, as another example, an alcoholic put in rehab and strapped to a bed until all the alcohol leaves his system. When he leaves the clinic he retains his sobriety. This person maintains his self-identity because merely having one's system cleansed from alcohol does not eliminate the fundamental attitudes associated with the condition. The recovering alcoholic will only retain his self-identity if he chooses sobriety despite its conflict with the previous fundamental attitudes and because of its particular characteristics. The only way these particular characteristics could have been chosen for

given the previous attitudes is if the person imagined how they *could* be valuable given a new set of fundamental attitudes and then adopted these attitudes.

Still, even under the Platonic Glasses mode of caring, it is a mystery how the actual revolution of preferences takes place. A person does not care along the new dimensions of an unprecedented self-synthesis until after making the self-subsuming decision to adopt them. Before adopting them, one cannot simply ponder them and then assign an amount of care to them because they are the very things that care springs from. Thus, they are distinctly phenomenal intentions and not the sorts of things that one can hold up hypothetically in a supposition.

One way the Platonic Glasses mode of caring will facilitate a self-revolution is if it creates a new concept that alters the mental landscape of the person thinking it. If this new concept alters the landscape *enough* or *in the right way*, then it may result in the re-ordering of its content along new fundamental attitudes; this is a self-revolution. There is basis for this suggestion in Nozick's discussion of how the closest relation schema structures concepts. He states that the overall informativeness of a particular classification using the schema depends upon what exists to classify. As the cases to classify change, judging closeness by some new relation R' might better constitute transitive identities in a time-slice space than did the old relation R. Likewise, tracing these transitive identities through time might, at some point, become more informative along a new relation of closeness (p. 57). Still, he maintains that the old concepts will retain some inertia because a prudent classifier will not change concepts of closeness every time the slightest digression occurs from the usual patterns in the objects he faces. The new classification must be more informative by a margin large enough to outweigh

the costs of changing, like the costs of changing habits, of becoming less compatible with current friends, and so on. "Yet", he claims, "when we move deep into the inertial realm, the old concepts will show more strain until there is a discontinuous shift in classification" (p. 57). This sort of radical break constitutes the full extent of a revolutionary self-synthesis.

III. Not all Self-Revolutions Preserve Personhood

The above considerations suggest that a change in fundamental desires is 'good' if it yields a more informative classification of one's mentality than did the previous classification. Perhaps a revolutionary self-synthesis is rational if and only if it results in a classification that is more informative than the last by enough of a margin to outweigh the costs of such a change of character. Still, it is unclear how a thinker is to know if the new desires will yield a more informative classification and why, exactly, a more informative classification is 'better' for the thinker than a less informative one. Even if thinking of a new concept can explain the revolution of preferences, the question arises as to whether this revolution has merit and whether the person undergoing the revolution deserves merit. The problem is that before the revolution one does not measure continuers along the *new* dimensions of the unprecedented self-synthesis, the appeal of which can only be understood under the new fundamental attitudes.

Nozick claims that a new reflexive self-synthesis justifies *itself* as progressive over the old by incorporating past self-conceptions into itself, attributing mistakes to them only if these mistakes are explainable (p. 91). However, his explanation implies certain consequences that clash with some very strong intuitions about personal identity.

When fundamental attitudes change, by the closest continuer theory, the borders separating internal and external histories of the self-identity that they constitute also change. The only sort of justification that can support explanations of 'mistakes' in a person's history is that relative to her current attitudes and the reflexively self-justified content of her current metric. Nothing in Nozick's account distinguishes a *real progress* between the pre- and post-revolutionary self from a *rewriting of history as if* it were a progress. A person that, under the programming of a cult, identifies as a member will view her non-member past as one of error. In reality, her non-member past was not in error, just 'not programmed'. A mentally ill person that identifies as Jesus may view his pre-Jesus self as self-identical but 'pre-revelation'. In actuality, his pre-Jesus self was not pre-revelation, but sane.

One option is just to accept that people cannot be wrong when they self-identify – there is no right and wrong about attitudes and the metrics they determine. Nozick verges on this view when he claims that there is a uniformity of delimitation achieved by the social matrix as it limits the possibility of one clumping oneself along any old set of relations synthesized around the act of reflexive self-referring (p. 107). He claims that the threat of punishment, institutionalization, or death will deter individuals from embarking on classifications that are dangerous or too weird (p. 108). However, this implies that if one lives in a dangerous or weird society, there are little or no limits (or arbitrary limits) on choices of metric within it. Strong intuitions indicate that self-syntheses can be regressive *in themselves*, not just for society, and that certain classifications are worth avoiding for their own sake.

I will describe two classifications that are worth avoiding for their own sake. First, what I call 'nonsense self-identifications', self-identifications as non-human objects or as other (dead or alive) humans. For example, 'I am a glass of orange juice'. This is nonsense because it is impossible for glasses of orange juice to self-identify. Other examples are 'I am Napoleon' or 'I am Stephen Harper'. Self-identification as Napoleon violates the strong intuition that Napoleon cannot self-identify because he is dead. On the other hand, Stephen Harper can self-identify, but if someone who is not Harper self-identifies as the entity that is Harper, this cannot be *its* self-identification because it is already self-identifying on Parliament Hill or 24 Sussex Drive. Nozick claims that in no possible world could something with an essential self-identity lack the capacity for reflexive self-reference because it undergoes synthesis by that very capacity and around it qua having it (pp. 90-91). A criterion for an object's self-identity, if that object lacks the capacity for self-reference, is in one sense arbitrary. Objects that maintain identity for a physicist, e.g. a hammer before and after it is broken down into its constituent particles, do not maintain their identity for, say, a carpenter. A grain of sand, for another example, has self-identity but it is not determinate, it is a vague concept and there is no straightforward answer, through changes to that grain of sand, as to when it stops being *that* grain of sand and starts being another or something else. Human identity is not relative to some particular purpose because it is established by self-synthesis. Since human self-identity is *intentional*, it is not relative to some particular aim like taking an accurate census. It can be its own aim and can thereby be intrinsic and essential. Still, under this construal, a person *could* self-identify, for example, as the *first* reflexively self-referring stone. The problem of induction is on this person's side. To rule out such

appeals, knowing they are probably false, it is the burden of the self-identity theorist to show where such a self-identification has gone awry on the inside, and the closest continuer theory, as it stands alone, does not do this.

The second sort of self-identifications that are worth avoiding in themselves are those that are detrimental to personhood itself and human flourishing, e.g. 'I am (essentially) a loser' or 'I am (essentially) a serial killer,' etc. As Nozick's theory of identity stands, there is no way of presenting these two sorts of 'mistakes' in self-identification from occurring. The verb, 'to be', is ambiguous in the English language meaning either predication e.g. 'John is sad' or identity e.g. 'John is the man with the sad look on his face'. In the case of *self*-identity and, particularly, self-identification, the verb 'to be' makes both an identity statement and a predication. Picking out one's best instantiated closest continuer is an act that not only means, 'I will be *that* person' but also implies, 'I will be *that* person *because* I will be a person that is strong (or sad, intelligent, Christian, etc.)'.

At first glance, these two impediments seem like very separate problems, e.g. a descriptive problem (self-identifying as something one simply *isn't*) and a normative problem (self-identifying as something one *shouldn't be*). However, I will argue that they stem from a common malfunction in that they are self-identifications that make one's identity less determinate. To overcome these two impediments, I show how marrying Nozick's theory of value to his theory of personal identity enables a distinction between value-increasing and value-decreasing self-syntheses. I term the former 'progressive' and the latter 'regressive'. Moreover, I show that the extent to which one's acts of self-identification make one's identity more valuable correlates positively with the extent to

which one is determinate. To elucidate the notion of a progressive revolutionary self-synthesis, I draw upon Nozick's theory of value in *Philosophical Explanations*.

IV. Value Theory

The purpose of this section is to give a coherent definition of value applicable to the current concerns about revolutionary self-syntheses. The next section will enter this notion of value into a model of human decision making. These two sections lay out the background information necessary for an explanation of progressive self-syntheses as those that increase the determinacy of identity so that a revolutionary self-synthesis, if progressive, will preserve personhood.

Progressive revolutionary self-syntheses should exclude both nonsense self-identifications and self-identifications that are detrimental to personhood. An account of progressive revolutionary self-synthesis requires, first, a definition of 'progressive', and second, an account of why it is rational, from the agent's perspective, to undergo a progressive self-synthesis. These two requirements are, at least at first glance, largely entangled. This is evident in Glaucon's challenge to Socrates, which Nozick addresses in his formulation of a theory of value. The challenge is: if one had a ring endowing the power to evade detection for any action one performs, is it in one's own self-interest to be moral? This problem is pertinent in moral philosophy because of the extent to which the demands of morality are, at least *prima facie*, contrary to the demands of rational self-interest. According to the theory of rational self-interest, it is rational to perform the actions most conducive to the satisfaction of one's most fundamental desires. The problem of showing how progressive self-revolutions preserve self-identity (and how

regressive revolutions break it) parallels the problem of Glaucon's challenge. It is easy to imagine cases where the actions most conducive to the satisfaction of a person's fundamental desires are not the actions that will be better for that person. If any unprecedented aspects of a self-synthesis result in an altogether different self-identity than the previous, then self-improvement is impossible. For example, the actions most conducive to a heavily addicted junky's most fundamental desires are not those that will get the junky off the drugs. Progressive revolutionary self-syntheses are required to explain certain cases of human improvement. I argue that in the cases where they are required, it is rational to perform them, even though they are not most conducive to the fundamental desires of the performing person. When what is better for a person involves a change in that person's most fundamental attitudes, a determinacy of identity over this change in attitudes is necessary if one is to say that personal progress in such cases is possible. I claim that maintaining and increasing the determinacy of one's self-identity is a deeper and more authoritative motivation than aims that are merely in one's rational self-interest.

This problem of reconciling a rationally self-interested with a 'good' self-synthesis extends also to cases of changes in attitudes that are not *most fundamental*. In cases where a change in attitudes is better for a person, we want to say that the person after the change is the same as the person before the change. However, we do not want to say that they are the same in virtue of what has not changed, that is, in virtue of the mere precedents in self-syntheses. For example, consider a man, Joe. Joe has a job in a mail room, lives alone with his cats, and is always getting down on himself and telling himself that he is no good. Today, however, is different. Joe receives a smile from an attractive

neighbour that inspires him to gain confidence. For the rest of the day, he walks with a jaunt in his step, smiles more, uncharacteristically joins a game of baseball on his way home and, moreover, hits a homerun! It seems unfair to Joe to say that he is the same person as yesterday only because he still has a nowhere job and lives with his cats. He is not the same person in virtue only of the facts in his life that have not changed. There is something about the change itself that expresses the potential he'd been carrying around with him. An expressed potential is the coming to light of internal properties of Joe that people (including Joe perhaps) didn't know were there.

The preservation of identity does not just consist in a lack of change but can also consist in change when it is the right kind. Joe's actions today do not show that old Joe died a little and a new Joe was born, but that there is more to Joe than has been meeting the eye—he has more of a personal identity than people (including Joe perhaps) thought. This is because Joe did *better* today than most days. Had he done worse, that is, had his actions today expressed less of a potential than his actions on most days do, e.g., he expends extra effort to avoid social contact, calls in sick, plays online slot machines all day, then this does not express a potential he'd been carrying around with him. In fact, this shows that what appeared to be the expression of potential may have been misinterpreted (even by Joe) as such. Would Joe really have rather been at home on the slot machines everyday when he was walking to work? If he had, then this is shallowness in Joe that perhaps his behaviour has been covering up and he has less of a self-identity than was thought. Thus, the current discussion is not only relevant to cases of full-fledged revolutionary self-synthesis but also applies to what, through the smaller changes in persons' lives, is responsible for the continuation of that person. I am arguing that it is not

merely the maintenance of precedent, but that the changes themselves can express the continuity (or lack thereof) of a person. Moreover, I will argue that some changes add to personhood by making someone a more determinate person than before while others diminish personhood by making someone a less determinate person than before. The current task is distinguishing between these two sorts of change.

Nozick distinguishes between moral pull and moral push. Moral pull is the characteristic in virtue of which situations place moral constraints on behaviour; moral push accounts for why and how a person can have motivation to behave in accordance with the constraints of moral pull (p. 401). There is an 'ethical gap' when the push is less than the pull (p. 401). 'Value' is the concept that connects pull and push, something's value is what pulls people toward it, perception of value is what pushes people toward it. Value is intrinsic if it is independent of whatever it leads to; it is instrumental if it is a function and measure of the intrinsic value it will lead to; it is originative if it introduces new instrumental or intrinsic value into the world; and it is contributory if its occurrence results in value that wouldn't have otherwise been there (pp. 311-313). This discussion will focus mostly upon intrinsic value, but I will mention originative and contributory value towards the end.

To explain how intrinsic value works, Nozick introduces the intrinsic value dimension D as an explanatory dimension accounting for the total value of anything by its score, including the score of its consequences, and of its encompassing wholes, along D (p. 414). He then goes on to specify conditions without which there could not be a measure across all types of value. The first is that value should be able to establish an

ordinal ranking of some sort over things, action systems, states of affairs, and so on (p. 429).

Next, he considers the function of value. He claims that value is something to which we are to have, when possible, a certain relationship (p. 429). By this, he means that there are certain appropriate responses to what values are like (p. 430). He considers a class of verbs, 'V-verbs', that specify these appropriate responses. They include 'care about', 'accept', 'support', 'affirm', 'encourage', 'protect', 'praise', 'be drawn toward', and 'nurture' (p. 429). The second condition upon D is that it can be a basic dimension of intrinsic value only if, "when some X is ranked highly along it, V-ing X also is ranked highly along it" (p. 430). Moreover, the third condition upon D is "if X is of high value along D then according to D anti-V-ing X [e.g. destroying, ignoring, hating, etc.] is of low or negative value" (p. 430).

Nozick claims that D either demarcates and yields a ranking of disvalue or that it has an opposite analogue that underlies and yields a ranking of disvalue (p. 431). The fourth condition is thus that anti-V-ing disvalue itself must have value according to D and the fifth condition is that V-ing disvalue itself has disvalue (p. 431).

The sixth condition upon D is that it be able to discriminate between the differing intensities of V-verbs and, for some value X, V-verbs that are more proportionate responses to X are of higher rank than those that are less proportionate to X (p. 431). The seventh condition requires that D ranks proportionate anti-V-ings of a disvalue higher than disproportionate ones (p. 431). The eighth condition is that a more intense anti-V-ing of a value is of greater disvalue than a less intense anti-V-ing of a value (p. 432). Ninth, there is greater disvalue to a more intense V-ing of a disvalue than to a less intense

V-ing of that same disvalue (p. 432). Likewise, holding the verb constant and varying the values: tenth, it is worse to V greater disvalues than lesser ones and, eleventh, it is worse to anti-V greater values than lesser ones (p. 432).

All things measurable along some particular value dimension, e.g. tastiness, beauty, pleasure, etc., constitute the realm of that dimension. Nozick claims that, across all these different realms, organic unity is the common strand to value (p. 418). An organic unity is a unified diversity where the measurement of diversity is relative to a set of dimensions along which things differ or are similar and the measurement of unity is relative to a set of unifying relations (p. 416). The idea is that organic unity satisfies the above conditions upon D for a larger realm of entities than any other characterization of D. Recall that it also affords the most informative classification in the process of entification as described at the beginning of this paper. Thus, organic unity has the potential to encompass what, at the end of the last section, appeared as irreconcilably prescriptive and descriptive problems concerning personal identity, the problem of nonsense self-identifications and of self-identifications detrimental to personal identity and human flourishing. In organic unity lies the potential for a common answer to the questions, 'who am I?' and 'who ought I to be?'

Nozick shows that organic unity satisfies the necessary conditions for D and that it satisfies them better than anything else does. The V-verbs are verbs of unification that establish and embody complex unities between persons and values. So where V-verb actions hold there will be positive degrees of organic unity (p. 432). For example, cherishing one's friendship establishes a greater organic unity between oneself and the friendship because it is an appreciation of the various aspects of the friendship and how

they come together to form something good, like the feelings of comradeship it evokes, how sharing one's favorite music (and other things one already likes) with this person intensifies its appreciation, etc. Likewise, anti-V-ing disvalue, he suggests, establishes unity one level up, at the metalevel (p. 433). Nozick does not elaborate on this point. It could occur if anti-V-ing disvalue serves as a protective belt, without which the organic unity of V-ing values would be lessened by corruption. For example, if threatened by enemies, thwarting their malintentions is a protective belt without which the organic unity of cherishing peace is impossible. Thus, organic unity is more than just stipulatively attached to D; rather, the meaning of organic unity as described here and in the beginning of this paper as the product of informative classification is the best characterization of D's content.

Still, it may not be that organic unity uniquely satisfies the conditions upon D or it may not be that there is a unique set of values that maximizes organic unity. Nozick formulates the further condition: "When C1…, Cn are the constitutive conditions… on value… then it is valuable (according to dimension D) that there be some dimension that satisfies these conditions" (p. 435). He points out that the philosopher's quest for objective values is based upon the value judgment that it is better that there be a unique set of objective values than otherwise (p. 434). On this condition, D must be valuable according to itself, it must be self-subsuming; this is a tight mode of organic unity, thus organic unity is valuable according to itself.

This last condition constitutes Nozick's "realizationism" about value. This is the view that we choose or determine that there are values that exist, but their character is independent of us (pp. 555-556). He points out that realizationism has been proposed in

mathematics, quantum mechanics, psychoanalysis, and literature, and thus is a coherent position and a possible structure for value theory (pp. 556-557). The view is that particular facts F do not by themselves entail particular value or ought statements. To do so they need as an additional premise the existentially qualified claim that there is *some* true value or ought statement (p. 567). Humans add this assumption and, through adding the assumption, justify it. Value, Nozick claims, does not need to already exist in order to come into being; it only needs to be possible that it would be a certain way if it were to exist. Once one can imagine how it would be, as outlining the conditions upon intrinsic value dimension D carried out, one can make it actual (p. 562). Nozick points out that it need not be better that there be value by any particular preexisting standard of value, but once there is value, its particular character is external to the value-seeker and, thereby, provides an external criterion for how to value it rightly and wrongly (pp. 563-564).

Nozick can still imagine other characterizations of D satisfying all of the above conditions, so they still might not uniquely specify organic unity. This is not a problem, he asserts, if one recalls the best instantiated realization mode of structuring a philosophical concept C, which holds that what C is depends upon what instantiated property or relation best fulfills the conditions associated with it (p. 442). Organic unity is the best instantiated realization of the conditions upon D and should another characterization of D better satisfy the conditions upon it, it would override organic unity, but until one is specified, organic unity is best instantiated. Moreover, should another set of values impose a higher degree of organic unity they would become the real objective values.

The purpose of this section was to characterize a standard by which to judge self-syntheses as progressive or regressive. The result is that a self-synthesis is progressive to the extent that it increases what we might call a person's overall degree of organic unity and it is regressive to the extent that it decreases a person's overall degree of organic unity. Admittedly, the precise sense in which a self-identity has a greater or lesser degree of organic unity remains unclear, but this will be developed throughout the remainder of the paper.

With this established definition and characterization of value, the next step is to demonstrate how it features in human decision making. Through an examination of Nozick's weight-bestowing model of decision making, it will become clear that value, as organic unity, has important implications for traditional conceptions about the utility of actions. On these traditional conceptions, the best act available to an agent is either the act with the best end state consequences (teleological) or is the act originating from the right sort of processes (deontological). The next section shows why Nozick believes that both of the traditional conceptions fail to capture an important characteristic of choices that can contribute their being the optimal choice. Nozick names this characteristic, "symbolic utility" and in the section after next I show that it can explain how progressive revolutionary self-syntheses preserve self-identity

V. Value and Decisions

The above external conditions upon what can characterize a dimension of intrinsic value do not capture the allure of value. Values themselves lack causal powers, but require value perceivers (pp. 436-437). This is how they serve as a bridge between ethical

push and ethical pull and it shows why progressive self-syntheses can be rational to perform.

Nozick specifies a structure that could explain how the perception of value results in ethical push. He formulates a weighing structure[6] that he calls the 'simple balancing structure'. It "...utilizes two open-ended lists of features of action: W (for wrong-making) and R (for right-making)" (p. 479). Plus, neither W nor R is empty so moral nihilism is false (p. 479). W and R are exclusive lists so no feature is in both (p. 479-480). W and R are not exhaustive so there are features of action that are in neither list and are morally neutral (p. 480). He outlines the following conditions upon inequalities between sets of features: First, they must be asymmetric: "if one set of features outweighs or overrides another, the other does not also outweigh or override it" (p. 481). They must be transitive: if A outweighs B and B outweighs C, then A outweighs C. Finally, adding more W features to an action cannot make it morally better and the more W features one adds to an action, the more morally worse it becomes (*ceteris paribus*) (p. 481).

By presuming that it is possible to assign numbers representing the differences between weights, Nozick formulates his 'principle 3': "It is impermissible to do act A if another action is available with less weighty W features such that the extra wrongness of A over that alternative overrides A's extra rightness over the alternative" (p. 488). This is the 'amended simple balancing structure', it recognizes that moral cost is unlike other sorts of cost because it does not simply maximize Ra > Wa; the W features of an action

[6] This is after he entertains and then rejects purely principle-based accounts (the 'maximization' and 'deductive' structures) because, first, there is the possibility of conflict between the supposedly exceptionless principles (p. 476). Second, they cannot explain how a person's judgment about a particular act often changes through learning more additional facts about the act so that he no longer judges it morally permissible or impermissible (p. 477). Third, it is simply the case that many people are unwilling to state or assent to any or very many exceptionless principles of determinate content (p. 477).

leave a 'moral residue', "Alternatives have to be sought, explanations and apologies have to be offered, amends have to be made…" (p. 489). An act's Rs or Ws have meanings independently of their mere proportions to one another.

To focus more precisely upon what is being weighed under the amended simple balancing structure, Nozick contrasts the teleological with deontological view of where, exactly, value lies. The teleological view concentrates on the best consequence while the deontological view concentrates on the right act. The teleologist speaks of the best act, Nozick claims, but the deontologist does not because he assumes that, "unlike 'right', the term 'best' applied to an action must match the best consequence" (p. 497). However, Nozick points out that the best act does not always need to be the act with the best end state consequences (p. 497). He claims that the acts available in choice situations involve different verbs that specify different verb-like relations to different end states, e.g. a bringing on, a preventing, a causing, etc. Since, on the view that organic unity is value, the different acts involve varying degrees of organic unity with other people and situations, their goodness does not need to vary directly with the goodness of their consequences (p. 497). He claims, "Doing that act with the highest score along the value dimension D is not the same as maximizing the D score of the world" (p. 497). The act which establishes the greatest organic unity between a person and the world is not necessarily the act that establishes the greatest organic unity in the world regardless of the role of the person. It follows from this 'person-bounded' perspective of organic unity that the W or R of an act is bound up with the identity of its performer and with the identity of the performers it affects. One reason why the W or R of an act leaves 'moral residue' could be that, should I perform a 'necessary evil', I wouldn't be the person I am unless I

recognized the evil and performed the appropriate ceremonies and recompenses to myself and others to maintain my self-identity in the face of it—even if these are merely feelings of regret, even if these feelings of regret are over something that was necessary.

The view of human self-betterment implied by the organic unity theory of value is neither strictly teleological nor deontological. Nozick entertains two models of betterment: the hierarchical lexical view, "If helping others or developing certain talents ranks higher than personal pleasure... personal pleasure takes second place to these" (p. 508), and harmonious development, "...your pleasure should be transformed and transfigured so that it now (also) comes from developing these talents or helping others" (p. 508). The theme of the transfiguration of pleasure so that it comes from the development of talents is present in the idea that self-progress depends upon self-syntheses that deviate from the precedent. The harmonious development view of self-betterment implies self-syntheses to the revolutionary side of the spectrum.

Harmonious development, Nozick points out, removes or drastically diminishes the divergence between self-interest and morality (p. 509). This is because, as persons develop a greater understanding of a thing in the world that they respond to, the actions they desire to perform in response to it change. The actions become more appropriate to the various characteristics that, unified, constitute it as the entity it is. The actions become more of a response to its organic unity, its value. Nozick states, "To be responsive to that valuable characteristic, therefore, is the way of being most responsive to that something, to the great diversity of it as pulled together in a unity" (p. 524). This explains why, under Nozick's construal, teleological accounts of decision making do not account for ethical pull, but the organic unity theory of value theory on the harmonious development

view of self-betterment does. On this theory, to look at the value of doing acts takes account of how one is responsive to them by linking up with them qua them and, just as importantly, qua oneself (p. 530). Attempting to maximize the world's value score, on the other hand, does not look at the character of one's relationship to things, but only at the total value everything ends up with (your relationship being just one of those things counted) (p. 530). The teleological approach does not account for moral pull. Focusing upon the degree of organic unity between each response to the world and the thing it responds to accounts for moral pull and it orders the responses differently than by the total sum of degrees of organic unity each response will ultimately introduce into the world.

The next section explains that the intention to create an organic unity can be motivated by its 'symbolic utility'. Actions satisfying organically unified intentions have high symbolic utility if they symbolize the other actions that, in different contexts and circumstances, would also serve these intentions. Thus, actions serving organically unified intentions serve intentions that span a variety of contexts and circumstances. Less organically unified intentions are less determinate across contexts and for this reason they justify a smaller variety of actions in a smaller number of situations. An action on a less organically unified intention symbolizes fewer other actions since fewer actions would express such an intention; thus, it has less symbolic utility. The next section takes up this theme, explaining symbolic utility in the context of progressive revolutionary self-syntheses.

VI. Degrees of Determinacy

The above section concluded that responding to the value of doing actions does not necessarily maximize the value in the world. This leaves room for an explanation of how a progressive revolutionary self-synthesis, a self-synthesis that increases one's value, is possible even though it cannot be the most rationally self-interested action on the pre-revolutionary preferences.[7] I claim that an action's utility is responsive to the degree of the determinacy of the identity of the person performing it. If the determinacy of a person's self-identity affects the utility of his actions, the action of identifying with a set of characteristics more valuable than the characteristics of prior identifications is explicable. Since, as I will argue, self-identities can be more or less determinate than one another, the utility of actions are commensurable over revolutionary self-syntheses. The notion of a progressive revolutionary self-synthesis cannot stand without such a commensurability because the action of identifying with the more valuable set of characteristics cannot itself be the most highly rated action according to the preferences of the prior identity; it is, itself, a new and different foundation of preferences. I will argue that an entire set of preferences, by the determinacy of the identity that holds them, can embody a higher (more valuable) understanding and appreciation of what value is than another set of preferences. I develop this argument in the next section. This section sets its stage by pinning down what it means to refer to an entity's determinacy and then establishing that self-identities can be of varying degrees of determinacy.

---

[7] This sense of the value of a self-identity is not one in virtue of which one can claim that all people are equal. One might say that all people are equal in terms of their potential to be valuable, but I am concerned with actual value. How a person can be more or less valuable, in this actual sense, will become apparent.

One way the performance of a revolutionary self-synthesis can override the most rationally self-interested actions serving the preferences of the prior identity is if these latter preferences stem from, as I explain shortly, a less determinate entity than that targeted by the Platonic glasses. A revolutionary self-synthesis on the Platonic Glasses mode of caring is doubly modal. Recall that the Platonic Glasses mode of caring targets the concept of what one would care about regardless of whether it is realized. So the Platonic Glasses mode of caring, when focused upon care itself, targets the concept of what one would care about caring about. It targets not what one would care about given the possible worlds in the actual world's neighbourhood of sufficient probability,[8] but, one level up, what one would care about caring about given this set of worlds. The question becomes, 'which cares or which conceptions of value are such that if I had them they would result in the highest potential value in the world that I could possibly realize?' If one's cares are such that one could realize a high potential of value, then they delineate a highly organically unified set of worlds. I will demonstrate that a self-synthesis around a type of caring delineating a more organically unified set of worlds will result in a more determinate entity than one occurring around a type of caring delineating a less organically unified set of worlds.

A self-synthesis that selects a type of caring that is more organically unified is a progressive self-synthesis. Moreover, the self-identity produced by a progressive self-synthesis is more 'real' in the sense that one can respond to it in a more nuanced, intricate, and delicate way than one could to any less organically unified set of characteristics identified with hitherto. I flesh out the 'reality' of an entity in terms of its

[8] This constitutes the worlds that, given one's beliefs and desires about and in the actual world, are likely enough to come about that their possibility features in one's plans and strategies for the future. They are, so to speak, such that one either takes out insurance against them or buys lotto tickets for them.

'determinacy'. The claim is that, *ceteris paribus*, if an action is in accordance with the preferences of a less determinate entity, then it will have less utility than that same action would if it were in accordance with the preferences of a more determinate entity.

By an entity's 'determinacy', I mean the extent to which it remains the same over changes to its constituent parts or to its environment. Entities, on this account, are more or less determinate; none are strictly determinate or indeterminate. However, this does not imply that there are degrees of identity; rather, it means that, to the extent that an entity is determinate, its classification as transversely and longitudinally identical (remember, this is what Nozick calls its 'entification') will remain the same despite changes to its parts or its environment, that is, despite changes in what there is to classify. Nozick's closest continuer schema of classification, in order to classify larger entities than 'atomic-point-instants', must sacrifice some of the similarity that is among what it groups together (p. 46). For this reason, he claims that his closest continuer criterion for identity "is the best Parmenides [for whom change is an illusion] can do in an almost Heraclitean [for whom all is flux] world" (p. 46). The best that Parmenides can do is to select the dimensions of sameness that give the most order to the flux, that best construe it as a manifestation of a constant underlying order. Of course, this underlying order must actually be there, that is, it must be there in virtue of certain recurring relational properties among the things it classifies. To the extent that an entification accomplishes such an order for the flux of constituent parts that constitute it over time, it creates a determinate entity.

An entity's determinacy lies in the extent to which the relevant criteria for its identity allow it to remain the same entity over changes to its parts and its environment.

The criteria for its identity are relevant if they actually do specify some similarity underlying the difference. One way they can do this is if they uniquely specify some *recurring* relational property that would not be otherwise specified. For example, Nozick claims that the relevant properties for a ship's identity are 'spatiotemporal continuity with continuity of parts' and 'being composed of the very same parts (in the same configuration)' (p. 33). These properties are relevant because, should one follow a ship around with them, one would refer to something in the world that is a thing in virtue of certain functional properties that one would not otherwise capture. On the other hand, if the criterion had been different and led one to refer to a pile of lumber that was once a ship as 'ship', then this scattering of the reference would discontinue its unique specification of 'a capacity for maneuverable flotation'. Had the criteria been different in this way for English-speaking people, the identity of the heap of wood with the floating structure would still have been a fact—it would just have been a more analytic and less informative one. If, again, only 'spatiotemporal continuity with continuity of parts' were relevant, then the pile of lumber *could be* identical to the ship, but the ship would lose some of its determinacy in that what a ship *is* has become more vague as it seems to no longer exist independently of the things that can constitute it. I am pointing out that, to the extent that an entity is determinate, the criteria for its identity pick out some functional property or properties. [9]

Since indeterminate entities are not constituted by the identity of some functional property or properties over time but just by whether or not certain other lower level properties are present, propositions about their identity accomplish nothing more than the

[9] An implication, then, of an entity's determinacy is that it is multiply realizable such that it can be realized by many distinct physical kinds.

convenience of not having to list all the properties they stand for. A heap of sand is not very determinate. One metric of sameness can classify two heaps as identical while other metrics can classify them as different, and the different metrics can all still pay equal tribute to the nature of a heap of sand. This is why the two heaps are not determinately the same or different. Comparatively, a watch has more determinacy because there are stricter conditions that must remain satisfied if it is to remain the same watch over time. For instance, one can slightly spread out the constituents of the heap of sand and to ask whether it is still that same heap is to ask a fuzzy question; however spreading out the parts of a watch will make it, determinately, no longer the same watch and even no longer a watch at all. Moreover, replacing constituents of the heap with new constituents that bear the same relation to the rest of the pieces of sand in the heap leaves the question open as to whether the heap is the same one. This is because the relations between the constituents of the heap are irrelevant; nothing about a heap implies anything about them. A person using a metric that classifies the watch with the substituted parts as a different watch would misunderstand what it is to be a watch because the function of a watch is more essential to its being a watch than is the material it is made of. If people went around tracing a watch's continuity based upon 'being composed of the very same parts', then if they came upon a little heap of scrap metal that used to be a watch they would wrongly believe it to still be one. The watch is a more determinate entity than the heap of sand because its identity depends upon the obtaining of more stringent conditions on the relations between its constituent parts than does the identity of the heap of sand. Metrics of closeness that respect these conditions pick out a functional property in the world, a property that can subsist despite changes to the material of its constituent parts so long as

they can subsist in the appropriate relations, and despite changes to its environment so long as they do not interfere with these relations.

I claim that an entity is determinate to the extent that it is organically unified and indeterminate to the extent that it is not. To the extent that an entity can undergo changes and remain the same entity, it is determinate; likewise with organic unity. Organic unity, by definition, is a whole that maintains itself in such a way that it maintains a certain independence from any conditions upon the holding of any of its parts. Nozick points out that something subject to few conditions on its parts is a whole: "...the more whole the whole is, the looser the limits are on which new parts might substitute in—for example, they need not be of the same material if they can serve the same function in the whole" (p. 103). Now, an organic unity, Nozick claims, "does something to maintain the integrity and continuance of the whole, unlike a heap which just lies there like a lump" (p. 100). Since an organically unified entity is a determinate entity, a personal identity that has a high degree of organic unity is determinate such that it can undergo great changes while remaining the same entity.

An organically unified entity's ability to maintain determinacy despite great changes to its environment contributes in two ways to the explanation of why organic unity is the best characterization of D as the intrinsic value dimension. First, it explains more particularly how a V-verb forms an organic unity with its object value. As pointed out above, valuing a value (with a V-verb) is a relation between a person and a value. Take the V-verb, 'caring', and the values, 'my future' or 'my hockey stick'. In both cases, caring characterizes a relationship between me and my value. Moreover, I claim, it requires a mode of mediation, a way in which it crosses between me and the things

valued. In the case of my hockey stick, the mediators involve hockey tape and wax, a safe storage place, my ability to see and feel the stick, the sinews and tissue that comprise these abilities, etc. – the mediators are potentially infinite. In the case of my future, the mediators involve my imagination, my powers of prediction, my savings account, the fridge full of food, marriage is a strong mediator for some people (it provides a guarantee that their future self will have a partner), etc. V-verbs maintain and increase the integrity and continuance of the relationship between persons and values insofar as they make the structure of this relationship an organically unified one; the function of valuing a value is realized to the extent that one can value it despite constraints on mediation. Ving a value is creating a determinate relationship between one's person and that value, a relationship withstanding changes in the constitution of whatever mediates the person's care to the value.

Second, the determinacy of an organically unified entity over changes explains the proportionality conditions between a V-verb and its object value. The action of Ving a value that is high on D should itself rank proportionally high on D. The V-verb's proportionality to the value consists in how hard it will struggle to maintain mediation with it. A constraint on the mediation of my valuing my future self is, for example, not having any money in my savings account or not being able to imagine where I'll be living. Since I highly care for (V-verb) my future (the value), I will struggle to acquire more mediation to it, e.g. direct my thoughts toward any vestige of it I can grasp or muster, work longer hours to *put* money into the savings account, and most importantly, exert effort in thinking of yet unexplored avenues of mediation to my future. An organically unified valuing of value holds despite constraints on mediation in the sense

that no particular mediator is necessary, it is only necessary that there *be* a mediator. This is to say that, to the extent that a V-verb is organically unified, it is 'multiply realizable' in that it can be realized by many distinct physical kinds. If I care about my future self, I will *find* a way to mediate this care. Since a hockey stick, on the other hand, is less valuable than my future, though I may use it as a road-hockey stick if it gets a crack, I will stop caring about it once it breaks because mediating my care for it by gluing it back together, or even keeping the pieces, is not worth the trouble. My care for the hockey stick is less multiply realizable in that I will only care for it if the appropriately convenient mediators are available.

As the organic unity of an entity is the extent to which it remains determinate withstanding changes in its environmental conditions and constitution, an organically unified self-identity also possesses this feature. The lesser the conditions upon the physical constitution of an entity's parts, the tighter the unity must be among its parts. Nozick conjectures that it is "the tightness of the unity relations which allows the looseness of the restriction on what may replace parts within a whole" (p. 103). Imagine a self-identity that depends upon the holding of certain of the diverse parts brought together in the self-synthesis—certain of these parts might themselves be only able to subsist under certain conditions. To the extent that a self-identity depends on conditions upon a part (or parts) it brings together, it is less of an organic unity. Moreover, the extent that a unified whole itself depends upon certain conditions for its subsistence also limits its degree of organic unity. If a unified whole depends on certain environmental conditions to remain a unified whole, then it is not a separate entity from them, but these conditions constitute a part of it (just like water constitutes a part of the human body). If the unified

whole can only be where these parts of it can be, then it is limited by these conditions upon its parts. The more organically unified a self-identity is the less subject it is to such conditions.

Now all personal identities are subject to some such conditions. For example, all personal identities depend upon sustenance from food and water because all living bodies depend on food and water and all personal identities (at least at this point in time) depend upon living bodies. Of more interest are the conditions to which some personal identities are subject but others are not. For example, a person whose self-synthesis includes friendship with certain particular individuals depends upon these individuals if she is to perform further syntheses on its precedent. If these friends abandon her, she will be forced to betray the precedent set by her previous self-synthesis and must perform a synthesis further toward the revolutionary side of the spectrum.[10] On the other hand, a person who synthesizes herself by including her capacity for friendship is not forced to abandon precedent if the particular friends she has *now* abandon her. The second person's identity is more organically unified than the first because the first is bound by keeping the same friends; the second can subsist without this condition. If the first changes friends, then her self-identity also must change; if the second changes friends, her self-identity remains invariant.

The situation of friendship could also fall under another category that is more typical in the case of love. If, say, a friendship with a particular person or the love of a particular person becomes valuable *enough* to me, then it will become worth it for me to identify with this particular friendship or this particular love. This is not a sacrifice of the

[10] She may be delusional, continuing to believe she has certain friends even though they have stopped being her friends. I don't address this possibility directly, but it falls under the purview of the next section.

determinacy of personal identity because in such cases the value of a particular friendship or love will be high, which means its determinacy will be correspondingly high. My proposition, then, is that if I am not a very determinate individual, then I will perform an unprecedented synthesis to identify with a love that is only mildly valuable—this identification will increase my determinacy; however, since it is only with a mildly valuable thing, my reliance on mediators of care for this thing will be proportionally high. On the other hand, if I am a very determinate individual, I will not identify with a love or a friendship unless it is *extremely* valuable and I will have a proportionately low reliance upon mediators to make it subsist. I will ensure it subsists despite changes in its constituent parts, e.g. the aging of its participants, changes in their character (that preserve their identity), etc., and despite changes in its environment, e.g. poverty, wartime, oppression (the stuff that great love stories are made of). In such cases, the utility of certain actions will be responsive to the determinacy of the love or friendship.

Since a self-identity with high organic unity does not depend upon any conditions upon its parts, when the Platonic Glasses mode of caring focuses upon a greater possibility of organic unity it focuses upon the possibility of an entity that can subsist and act in more places and under more different conditions than the self identity from which it springs. The ability to subsist and act in as many different places and under as many different conditions as possible is explicable in terms of the *span* and *intensity* of executable intentions. Executable intentions are intentions that could be satisfied with a reasonable degree of effort. For example, my intention to go to sleep early tonight is executable but if I had the intention to climb the highest mountain on Jupiter it would not be. *Span* is the variety of things with which one can engage in complex relationships e.g.

the Renaissance man has great span. *Intensity* is the complexity of the relationship constituted by any given engagement, e.g. the relationship between a programmer and a computer has great intensity, the relationship between a retailer and a cash register has less.

A set of executable intentions with great intensity results in an organic unity of potential responses to stimuli corresponding to the organic unity of one's area of expertise—I call this a capacity. The astro-physicist's study is intense because she is extremely *capable* in her area of study (if she is good at her work). She is able to engage complex phenomena in very tight organic unities with her mental processes, instruments, and colleagues. There is an organic unity both in the world (because she perceives complex phenomena and can predict its consequences e.g. the orbit of a moon) and in her mind (because she can construct complex scenarios e.g. the landing of a spacecraft). The organic unity of the relationship between her and the world enables the significance of her perceptions and the power of her constructions. Her study creates span, e.g. what as a child she saw as blobs in the sky she now sees as solar systems, but she does not start out with span because the time and effort taken to acquire her skills probably requires sacrifice in other areas like athletics, music, mechanics, and horse-riding. As a general rule, the more capacities for weaving organic unities in and with the world that a person takes on, the less intense any one of these organic unities is apt to be. A progressive self-synthesis that increases a person's capacity through increasing the span and intensity of his intentions increases the determinacy of his identity.

To the extent that a person's fundamental attitudes, and the sub-attitudes they envelope,[11] enable a high span and intensity of executable intention, this person's identity is more *real*. [12] Since personal identity is a theoretical entity, not an entity that can be directly observed, but one that is hypothesized to explain the underlying unity of both our own and others' thoughts and behaviour, my criterion for a self-identity's determinacy is its *robustness*. A robust concept or law, on William Wimsatt's interpretation, has multiple-connectedness within a theoretical structure to the observational results on its periphery. Multiple-connectedness is what determines the meaning of Hilary Putnam's (1962) 'theoretical definitions,' "multiply connected law-cluster concepts, whose meaning is determined by this multiplicity of connections" (2007 p. 55). These are opposed to the trivial, because not multiply connected, analytic definitions that serve as laws not for natural kinds, but for synthetic classes, like bachelors, grouped together only in accordance with a single legal aspect - such laws are preserved come what may with few repercussions (p. 55).

When persons of regressive self-syntheses refer to their 'self', they verge on analytic definitions of this term because the sorts of values they hold hardly extend beyond their immediate context. For example, when a person that defines himself only by a brand of beer and watching football says 'I', he doesn't refer to a great many potential situations beyond those he is ever currently in; moreover, since his relationship with his objects of care is not very organically unified, a relatively small change in environment or available mediators of care will easily trigger a revolutionary self-synthesis. This person's self-identity verges on analytic because it is little more than a stipulated term for

[11] For example, my care for my family envelopes my care for my brothers, my mother, my father, my grandparents, and so on.
[12] Nozick develops this line in his book, *The Examined Life*, 1990.

the sorts of actions in which he is partaking at the place and time he utters the word 'I'. On the other hand, when persons of progressive self-syntheses refer to their 'self', since their cares and capacities out-span any immediate context, but hold for a great many human situations, the term is more theoretically adequate in that it *does* more by actually synthesizing a much larger number of situations in which action is possible.

Perhaps some will find it strange that capacities characterize a person's identity. Intuitively, one might claim, it is a sort of 'internality' to people that constitutes their identity. However, capacity, in the sense I am using it, constitutes internality. Note that it also connotes internality in the ordinary sense of the word. According to the Oxford Dictionary, 'Capacity' has two meanings: 'the ability or power to do something' (the one I have been using) and 'the maximum amount that something can contain or produce' (the one I am now about to claim for my use of the word).[13]

If organic unity is to explain what a person is as constituted through self-synthesis, then it cannot be *itself* internal to a person because this presupposes the person (and thus does not explain him or her). Rather, it constitutes a person to the extent that it *is* the internal to that person. The more different places weighed by the complexity of the ways that the organically unified person can engage the world is the degree to which there is an 'internal' to that person. For one thing, capacities remain capacities whether or not they are in use, e.g. I have the capacity to make bank shots on a pool table even though I'm not doing it right now. This shows that capacities can travel around e.g. my capacity to play pool is here right now in front of this computer. There is another reason capacity can constitute internality. If creative activity involves applying skills developed

[13] http://www.askoxford.com/concise_oed/capacity?view=uk

in response to one domain of experience to an entirely new domain of experience, this explains why creativity is commonly considered an apt expression of individuality. Moreover, the capacities of a person are the choices that person has in any given situation; they are, under any metaphysical construal you please, a person's freedom. They are his repertoire of responses both to his environment and to his own self (as in my capacity to think about my own thoughts). Lastly, this view points toward an explanation of the particular phenomenology of being a certain person. It 'feels' this way to be me, to think these thoughts of mine, to miss my girlfriend, because these situations are parts of larger systems of responding to stimuli that constitute my capacities in the particular way, and in association with the particular reasons for which, they have developed. The particular feeling of each response depends, largely, upon the responses with which it has been associated and the strength of these associations. Some might argue that there is no 'feel' to being me because, after all, I cannot compare this to the feeling of being anyone else. What I mean by 'how it feels to be me' is 'how it feels to feel things in precisely the way that I do'. Many will at least accede that people can feel the same thing different ways, e.g. some people like cold showers and others do not. This can explain how differences in such feelings correspond to differences in personal identity.

VII. The Rational Self-Revolution

The utility of actions is undermined to the extent that the preferences they serve belong to an indeterminate entity. I claim that actions meant to satisfy the preferences of a less determinate identity have less utility than would similar actions meant to satisfy the preferences of a more determinate identity. Ordinal rankings of preferences usually occur only against an invariant backdrop of fundamental attitudes. If a self-identity can

transcend a change in fundamental attitudes, as I have been suggesting, then this 'backdrop' to preference rankings becomes another variable affecting their ranking so that they are ranked not just relative to one another but relative to one another on different backdrops. The evidence that the utilities of actions *do* respond in this way to changes in the attitudinal backdrop of the preferences they serve is that there are fluctuations in the utility of actions best explained as responsive to the degree of a self-identity's determinacy under a given set of fundamental attitudes. Nozick's notion of the symbolic utility of an action can capture this fluctuation. I claim that fluctuations in the symbolic utility of a revolutionary self-synthesis (or of syntheses insofar as they are not based completely on precedent) is best explained as responsive to the relationship between the degree of the determinacy of the pre-revolutionary self-identity and the degree of the determinacy of the potential post-revolutionary self-identity.

Nozick characterizes an aspect of utility that he calls 'symbolic utility.' He develops this notion against a consideration of Newcomb's problem.[14] Two different principles suggest two different courses of action in the problem. The *evidential expected utility* of an act *A EEU(A)* "...specifies the expected utility as weighted not by the simple probabilities of the outcomes but by the conditional probabilities of the outcomes given the actions" (1993 p. 43). Since the evidence suggests the predictor will predict your own actions correctly, this principle suggests that you pick the single box. The *causally*

[14] Nozick gives this brief description of the problem: "A being in whose power to predict your choices correctly you have great confidence is going to predict your choice in the following situation. There are two boxes, B1 and B2. Box B1 contains $1,000; box B2 contains either $1, 000, 000 ($M) or nothing. You have a choice between two actions: (1) taking what is in both boxes; (2) taking only what is in the second box. Furthermore, you know, and the being knows you know, and so on, that if the being predicts you will take what is in both boxes, he does not put the $M in the second box; if the being predicts you will take only what is in the second box he does put the $M in the second box. First the being makes his prediction; then he puts the $M in the second box or not, according to his prediction; then you make you choice" (1981 p. 41)

expected utility of an act *A* *CEU*(*A*) specifies "…a probability relating the outcome to the action, this time not simply the conditional probability... but some causal-probabilistic relation indicating direct causal influence" (p. 43). Since the only causally relevant variable is one's own action, and since one's action cannot affect whether the money is in the second box, this principle suggests you pick both boxes. Nozick notes that, though people tend to operate on one or the other of these principles for any particular description of the problem, varying the amount of money in the first box can get them to change principles. Those who initially chose both boxes are unwilling to continue to do so if the amount of money in the first box is lowered to $1; people who initially chose only the second box will desire to choose both boxes if the amount of money in the first box is raised to $900,000 (p. 44). So Nozick claims that associated with each act is a decision value *DV* that is a weighted value of its *EEU* and its *CEU*. These weights are assigned to a principle proportionally to a person's confidence that that principle will guide her rightly (p. 45).

Now, Nozick suggests that further decision-making principles can be added to this weighing structure. He suggests adding symbolic utility, which incorporates the outcomes and actions symbolized by the act (p. 47). Evidence for symbolic utility is the persistence of an action in the face of strong indications that the action does not actually have the presumed causal consequence. Nozick's example is that antidrug enforcement measures *symbolize* reducing the amount of drug use (p. 27). Symbolic utility is the utility an act has independently of the utility of its expected consequences. Just as the *CEU* of an action is determined by its causal-probabilistic connection with the standard utility,

symbolic utility is constituted by a different and unique kind of connection to the standard kind of utility, a symbolic connection (p. 48).

Nozick claims that symbolic utility connects to standard utility in something like the way that metaphorical meaning connects to literal meaning. An act's symbolic utility depends upon what other acts are available, upon *their* payoffs, and likewise upon the acts and payoffs available to other parties—what an act symbolizes can differ from particular situation to particular situation (p. 55). For example, in the Prisoner's Dilemma,[15] the symbolic utility of an action will change from play to play depending upon the actions of the other party. For example, doing the cooperative action may symbolize being a cooperative person more and more as the other party continues to defect (p. 59). However, defecting in the prisoner's dilemma can also have symbolic utility, e.g. it could symbolize 'being rational and not being swayed by sentimentality' (p. 57). Nozick claims that *DV* leaves room for how 'what sort of person a person wishes to be' can feature as an explanation of their choices. He claims that, while psychological explanations about personal identity traditionally served as an excuse about why people *deviate* from rationality, now they can serve as a component *within* the rational procedure of decision (p. 57).

Symbolic utility works like this: "The action (or one of its outcomes) symbolizes a certain situation, and the utility of this symbolized situation is imputed back, through the symbolic connection, to the action itself" (p. 27). The standard utilities of certain

[15] Nozick gives this brief description of the dilemma: "a sheriff offers each of two imprisoned persons awaiting trial the following options. (The situation is symmetrical between the prisoners; they cannot communicate to coordinate their actions in response to the sheriff's offer or, if they can, they have no means to enforce any agreement they might reach.) If one prisoner confesses and the other does not, the first does not go to jail and the second will receive a twelve-year sentence; if both confess, each receives a ten-year prison sentence; if both do not confess, each receives a two-year sentence" (p. 50).

(symbolically connected) situations are 'imputed' along a symbolic connection resulting in symbolic utility as the standard utilities of certain (causally connected) situations are imputed along a causal connection resulting in causal expected utility. Imputation is like conduction; as only certain types of materials can conduct electricity, only certain types of relatedness can conduct utility. Symbolic utility is one type of relatedness that can conduct utility; a person's identity is integrally wrapped up with where (and to what) symbolic utility gets conducted for that person.

One way a person's self-identity determines the symbolic utility of her actions is through principles. Nozick points out that as principles serve the *inter*personal function of providing reliable bases for predicting the actions of others they serve the *intra*personal function of allowing us to rely upon our future self so long as we can cause its behavior to stem from similar principles as our present behavior (p. 14). Now, an action on principle can *symbolize* all the other actions that *would* satisfy the principle in different circumstances (p. 26). An action performed on a particular principle, he claims, can symbolize all the other actions that principle justifies (p. 26).[16] For example, a single action of voting probably will never make a difference to which party wins an election but symbolizes the other actions justified by the principle that one ought to vote and has imputed to it the utility of these actions. The utility of all of these actions together, imputed back through the symbolic connection to the act of voting, explains why voting is worth anybody's time.

[16] As principled actions are symbolic, Nozick proposes that the act of acting on principle is symbolic, symbolic of rationality (p. 40). Nozick characterizes rationality as a matter of reasons and reliability: we use reasons because they most often allow us to arrive at true belief, thus by using reasons we make our belief-forming processes reliable (p. 67).

The symbolic utility of a certain situation can fluctuate for a person and it does so when the person himself changes. Nozick uses the example of a person who has taken steps to reduce his guilt for whom the utility of being guilt-free has thus become less (because there is now less to deal with) (p. 28). As this person identifies less with guilt in his self-syntheses, the symbolic utility of his guilt-reducing actions, e.g. giving to charity, becomes less.

One might object that what we've been calling 'symbolic utility' could be explained away as just the simple expected utility of doing an action for the motive behind it. For example, perhaps the principle that everyone ought to vote motivates my act to vote and my preferences assign a high utility to the actions motivated by that principle. However, this does not explain *why* I assign a high utility to actions motivated by *this* principle. This is ok on what Derek Parfit calls 'desire-based theories' according to which practical reasons, our reasons for acting, are based only upon certain facts about what would best fulfill our present desires (forthcoming, p. 30). He rejects this type of theory, however, because it implies that people cannot have reasons for their ultimate aims—nothing matters, he claims, on this theory (p. 44). Instead, he opts for a value-based theory in accord with which we have reasons for the desires or aims that our acts will fulfill, reasons that are not *just* these desire or aims, and we would have these reasons even if we did not have these desires or aims (p. 30). If we adopt the value-based theory, then to say an action's utility is just the utility of performing it for a certain motive is not to explain its utility at all. Rather, there are reasons for assigning high utility to actions for certain motives and it is not the assigning utility that distinguishes these motives as important but the reasons for doing so. Symbolic utility does not merely identify these

reasons with a person's most inexplicable desires, but locates them in the person's conception of the value of particular actions where a person's conception of value can ultimately reside in beliefs about the sorts of things that are valuable.

According to Nozick's realizationism elucidated above, a person's choice that there be value in the world is a condition for its existence but its character is then independent of the person that so chose. The utility of voting is high because, were the principle that everyone should vote realized, the utility of all the actions satisfying this principle would be high for someone who values democratic representation, and this (or some of this) 'would-be utility' gets imputed as 'actual utility' to the singular act. For such a person, the choice that there be value, that there be an intrinsic value dimension D, occurred in his self-synthesis, but once the self-synthesis has established his self-identity, the value of democratic representation from which his act of voting derives its symbolic utility is a fact about the world and not an attitude of his toward the world.

Another possible problem with this conception of symbolic utility is a worry of rule-utilitarianism. If the action of voting is compatible with a plethora of different principles, then which principle determines the other actions to which the act of voting symbolically connects? If nothing pegs down a particular principle, the symbolic utility of an action is indeterminate. The answer is that the principle determining the symbolic utility of an action is the principle in accord with which the subject intends to perform that action. The aggregate subjective utility of all actions that would be performed in accordance with the successfully applied principle, given the agent's values, gets conducted back to the agent's action along their symbolic connection to it. This does not mean that the agent makes, by any objective standard, the *right* decision whenever an

act's decision value is high due to its symbolic utility. For example, even though the American Government's 'war on drugs' might have high symbolic utility given the values of the administration, perhaps it is a drastic waste of tax payers' money and wrong because it infringes upon certain rights and freedoms that are, for the sake of argument, more justified than the values of the administration. The symbolic utility of an action is not an infallible indicator of its objective value, but it can motivate the action's performance in situations where it is the most valuable (organically unified) action but not the one that yields the highest expected utility. This is important because one such situation is that of a progressive revolutionary self-synthesis.

My suggestion is that the action of undergoing a progressive revolutionary self-synthesis has the symbolic utility imputed back to it from all the intentions that *could* be satisfied if one *were* to undergo the self-synthesis. A self-synthesis is progressive if it has high symbolic utility because it reaffirms and strengthens one's conviction that there is to be value. The strengthening of the conviction that there is to be value will have high symbolic utility because a self-identity more committed to the existence of value will have a more wide-reaching and interconnecting network of possible symbolic connections serving the distribution of value to particular actions than one that is less committed. For the more committed self-identity, there is an increase in the flow and regulation of symbolic utility because value will imbue more things than it did before, that is, the intrinsic value dimension D will have a more inclusive realm than it did before. Thus, one will be motivated to perform a greater quantity of valuable actions than before.

This increase in value cannot occur by a mere relaxing of the conditions upon value so that any old thing becomes valuable. This will not work. One cannot say that 'X is valuable if it has extension' and thereby include all things; this will actually leave out *a lot*; colours will not be valuable for instance, nor any of the characteristics that could distinguish things with identical extensions, nor relations or functions. Rather, in order for all things *to be able to be* valuable, fewer things will be valuable in themselves. Perhaps it is sufficient to say that value is organic unity for theoretical purposes, but we might not have given a substantive enough characterization of organic unity to serve the functions of a personal identity (one of these to facilitate the creation of value and/or the contribution of value to the world). Certain sorts of organic unity simply cannot be e.g. circuits cannot be created from non-conducting material. For self-identity, an understanding of the types of organic unity there can (and cannot) be will provide a flow of symbolic utility more conducive to the creation of value than an uninformed commitment to organic unity of value.

The symbolic utility of undergoing a progressive self-revolution comes into play through viewing, through Platonic glasses, a different set of fundamental attitudes than those one currently has, including the attitude toward the existence of value, and the actions that would be in accord with them, which they symbolize. A set of fundamental attitudes, the adoption of which has a high degree of symbolic utility because of a strengthening of the attitude toward the existence of value, has what I call 'much capacity'. If a person has an intention that she could satisfy if she wanted to, then her ability to satisfy it is a capacity.

Performing a self-synthesis strengthening one's attitude toward the existence of value constitutes an increase in the capacity of one's self-identity. As the last section showed, the determinacy of a self-identity is a direct function of its capacity. Thus, performing a self-synthesis strengthening one's attitude toward the existence of value constitutes a corresponding increase in the determinacy of one's self-identity. The action of becoming a more determinate person symbolically connects to the actions falling under the greater capacity of this more determinate person and the utility of these actions are imputed back to the performance of the revolutionary self-synthesis. Likewise, this explains the utility of performing those self-syntheses lower on the revolutionary scale but still not completely on precedent.

It may seem appalling at first that I am suggesting that some people have more warrant in caring for the continuation of their self-identity than others. I claim that the more determinate a person's identity is, the more rational it is to care about its preservation. Perhaps it is of comfort to hold that all people have warrant to the second-order care, desiring to desire to continue themselves (or desiring to value themselves). This account reconciles what would otherwise be a conflict between the beliefs that all people are in some way equal (have equal rights) and that it is good for all people to better themselves. If nothing I do can make me more valuable than others, or even more valuable than myself yesterday, then why should I better myself? On this account, though I am equal to others and all stages of myself in the respect that I *could* be as valuable as anyone else, I have motivation to better myself in that I am not actually as valuable as I could be.

The determinacy of one's self-identity is the extent to which 'that which continues' is an entity onto itself distinguishable from other entities. A personal identity that has a high degree of organic unity is robust in that it has more continuity through time than one that has a lower degree of organic unity, and one with a high degree of organic unity attaches more weight to continuing oneself through time than does one with a low degree of organic unity. I address this in the context of an objection from Derek Parfit in the next section.

VIII. Reductionism and the Determinacy of Identity

I have argued that, to the extent that one's identity is more valuable, one can care more for its continuation and thereby make it more determinate. Derek Parfit objects to the idea that a person's self-identity can ever be determinate and would object to the idea that the Platonic glasses mode of caring features in the determination of one's self-identity. This is because he is a reductionist about personal identity, which means that he holds, first, that a person's identity over time consists just in the holding of certain more particular facts; and second, that it is possible to give an impersonal description of these facts, without reference to or presupposition of the person involved (1984 p. 210). Moreover, he claims that reductionism about some sorts of things sometimes implies that their identity is indeterminate (p. 213). The sorts of things in question are those captured by vague concepts, like 'heap' in 'heap of sand', and he thinks that personhood is one of them (p. 232).

Parfit claims that personal identity is one of these vague concepts when, for instance, he insists that there can be no branching in psychological continuity. Branching

cases occur where there are ties between continuers, e.g. the Star Trek transporter case above. In such cases, Parfit holds, the concept of personal identity breaks down and asking who the original person continues as is to ask an 'empty question' that has no answer (p. 235). This does not mean that it describes different possibilities, any one of which might be true where one of them must be true; rather, it gives different descriptions of the same outcome, the truth about which we could know without choosing any particular description (p. 260). The question about personal identity, he claims, is not a matter of the best representation of the facts, but about how to best regard these facts; it is not a theoretical but a practical question (p. 260). On this view, to ask for the truth or falsity of a statement about personal identity is to misunderstand it. For Parfit, personal identity is nothing more than a useful concept relative to some further end, e.g. making a census, because it makes quick reference to entities that are more or less like others (stages of a person), but are not identical in any intrinsic or essential sense. I argue, on the other hand, that there are some cases where personal identity is determinate, where assertions of a person's identity do, at least to some extent, have a truth value. By 'at least to some extent', I mean that the assertions can be truer than false or more false than true, but not indeterminate with respect to truth or falsity.

It is on the belief in the indeterminacy of the truth-values of *all* statements about personal identity that Parfit's view is incompatible with the Platonic Glasses mode of caring and, more generally, with self-synthesis as an application of the closest continuer theory, which allows for the possibility of the determinacy of identity. Nozick writes, "I do not use the [closest continuer] schema to identify something as an aid to my goals; I use it to identify whose goals are mine" (p. 61). A self-synthesis on the Platonic Glasses

mode of caring transforms one's values into a unified representable fact that *is* oneself. If, on this view, you synthesize yourself as a valuable entity, and thus a determinate entity, and go on to care about oneself as a determinate entity, then you are one and it is true that you are one.

Parfit argues that not only is personhood a vague and indeterminate concept that catches nothing essential about human beings, but that it does not even locate what matters in situations that exploit its vagueness (like ties between continuers and self-revolutions). To set the framework for this argument, he asks us to consider all the possible degrees of psychological continuity (the 'psychological spectrum') and all the possible degrees of physical continuity (the 'physical spectrum') (p. 231). He defines psychological continuity as the holding of overlapping chains of strong connectedness. There is enough connectedness for personal identity "if the number of direct connections, over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person" (p. 206), presuming such things can be counted and compared. A direct connection occurs when a memory, a thought, a feeling, etc. connects directly with a past intentional object. Continuity is a chain of overlapping direct connections, e.g. if thought B is about thought A and thought C is about A, and if A, B, and C are successive, then C is continuous with A (p. 205). He defines physical continuity as the continuing to exist of enough of a person's brain so that it remains the brain of a living person as long as there is no branching in its continuity (p. 208). He claims that there is much reason to believe the carrier of psychological continuity is the brain, and that psychological connectedness could hold to any reduced degree (p. 238). Put the psychological and physical spectrums together and one gets 'the combined spectrum' (p. 236).

To demonstrate that personal identity is not what matters in situations exploiting the vagueness of 'personhood', Parfit engages us in a thought experiment about the combined spectrum. The point of the experiment is to demonstrate that personal identity is not what really matters in situations where it is hard to discern whether I am still myself (like branching cases). At one end of the combined spectrum, I remain completely physically and psychologically continuous with myself; at the other, scientists will replace my brain and my memories, bit by bit, with Greta Garbo's brain and her memories. At the various degrees in between the extremes of 'before any surgery' and 'after the complete reconstruction', the scientist replaces proportions of my body (most importantly, my brain) with proportions of Greta Garbo's, e.g. one cell, one thousand cells, all cells but one, and many steps in between. Parfit argues there is no determinate answer as to when I cease to be myself and begin to be someone else (p. 233). Thus, he claims, what matters is relation R, not personal identity. Relation R is *enough* psychological connectedness and/or continuity as defined above in terms of direct connections (however these connections come about) (p. 313).

Parfit does not believe that revolutionary self-syntheses preserve anything that matters because Relation R does not hold between the pre- and post-revolutionary person; therefore, he would claim that calling the pre- and post-revolutionary person 'self-identical' achieves nothing. A deep change in attitude, such as that of a self-revolution, will violate R by cutting most, if not all, direct psychological connections including the connections maintaining even the most ordinary beliefs and memories, like the preference for sour over sweet tastes, for the blues over jazz, for blue over green, etc., and for all the particular instances that these preferences trickle down to determine, e.g., for the blue t-

shirt over the green one. Take a person whose majority of memories connect directly to past intentional objects as do the majority of his other psychological features, e.g. beliefs, desires, etc. If this person undergoes a deep change in attitude, then most of these psychological connections will no longer be direct. People remember the things they do and in the way that they do because of how they feel about these things. Of my childhood, for example, I remember with the greatest vividness those things I liked and hated the most—attitudes guide memories and, moreover, affect their content. The things I hated I remember *as* ugly, e.g. a certain bully's scowling face and the menacing tones of his voice. People with different attitudes toward him remember him differently e.g. his mother might remember his smiling face and his playfulness; the things I liked I remember as good e.g. I remember games of capture-the-flag in the neighbourhood *as* fun. One of the reasons flag might be described as fun is because its memory exists in a network of memories. It is a memory closely related to the thrill of beating an opponent, of hiding and sneaking, of pretending to be a soldier; likewise, the memory of the bully is closely related to the memory of being bullied. In these cases, the content of a memory is good or bad derivatively because of the goodness or badness of the memories it is associated with. Even so, the reason these memories are in *these* networks, and not in some other network, is because of the way I *do* feel about the memories they are related to. This would apply even to the most common sorts of memories, like that I prefer calm days to stormy days or that I dislike mince-meat pie. Should my attitude change to that more like Heathcliff's and less like Catherine's, as in Bronte's novel, *Wuthering Heights*, then I would like stormy days more. Likewise, parents trick their children into liking their food all the time by altering their attitude toward it, telling them it is a 'different kind'

when it is not or preparing it in a fun way, e.g. making statues out of the mashed potatoes. Should my fundamental attitudes change, the networks of my memories would change shape and this would redistribute the derivative meanings of memories for all those memories right down to those considered most trivial and distant from questions of personal identity, e.g., the preference for pie over cake.

One might object that a change in my current attitudes need not affect the networks of my memories because the networks of my memories are shaped by the attitudes I was having at the time I was storing them e.g. I could remember having fun even though I no longer have fun but am now bored all the time. However, even this makes the memory of fun indirect, as it must get filtered through my boredom, which casts fun through a different interpretation than it otherwise would receive. Boredom, for example, gives it a nostalgic edge; if I were not bored now but *extremely* happy, the memory might not even seem fun anymore but drab in comparison. This response to the objection is reinforced by the point that not all the attitudes attached to my mental states can be derivative; otherwise, they'd have nowhere to derivate *from*. The non-derivative attitudes are my current most fundamental attitudes towards things. The old networks of memories are not independent from the newer networks but embedded in them. A word partially takes on its meaning from the sentence in which it occurs and this sentence partially takes on its meaning from the paragraph in which it occurs. Memories, I claim, take on their meanings in a fashion analogous to this.

Although relation R is not preserved over fundamental changes in attitude, personal identity of these changes is both a fact and is what matters if three conditions obtain. First, the fundamental change in attitude must occur via a self-synthesis on the

Platonic glasses mode of caring; this is the theoretical condition. I demonstrated this in section two by showing that a revolutionary self-synthesis cannot occur on the Best Realization mode of caring but can only occur through the Platonic glasses mode. Second, the self-synthesis must constitute an increase in organic unity; this is the practical condition. The last section demonstrated that the determinacy of a self-identity correlates with its organic unity. Thus, if a revolutionary self-synthesis decreases the organic unity of a self-identity, this identity is to that extent *less* determinate and thus it is appropriate to say that less of it is preserved because it is less of an 'it'. This constitutes a reason why personal identity matters because if personal identity is not preserved over such changes, personal progress is impossible. First, radical instances of personal change will never be considered progress because they will always result in an entirely new person than before and therefore never in a better person. Second, and more importantly, as in the instance of poor Joe above, the consequence will trickle down to all instances of progress such that whenever a person progresses he will never be the same person in virtue of his progress but always in virtue of the aspects of personhood that have not progressed.

Still, however, a third condition is required because I can imagine the case of a revolutionary self-synthesis into a more organically unified entity that yet, intuitively, does not preserve self-identity. I am claiming that Greta Garbo cannot be me. Now, it is a vague question, to say the least, to ask if *she* is more organically unified than I. So let's say I *know* that Mohandas Ghandi's self-identity was more organically unified than my own. Now I have already said, in Section III, why it is nonsense to self-identify as a pre-existing person. So why does it appear now, from the inside, that there is nothing wrong

with performing such a self-identification? Just because a personal identity is more valuable than another and at a later stage in time than that other it does not follow that it is the same personal identity; something further is required. The problem is that nothing grounds the practical condition in the theoretical condition.

The third condition is that the increase in organic unity occurs *because* it is an increase in organic unity. And this, I claim, can only occur if the content of the new (progressive) metric is chosen *because* of the content of the old (regressive) metric. Recall that a new most informative classification occurs when new fundamental principles of ordering produce a more informative classification than the former principles. In the case of self-identity, it is the self-synthesis that performs the classification in accordance with which I am myself; a person's fundamental attitudes are the principles of this classification. My current fundamental attitudes yield the most informative classification if they order my priorities and beliefs, given my environment, so that my life has the most organic unity. In a revolutionary self-synthesis from old fundamental attitudes to new ones, the former attitudes might have *initially* been producing an informative classification, but then begun to produce one less informative than other possible classifications. This situation provides the opportunity for new fundamental attitudes to usurp the old by realizing one of the classifications that would be more informative.

The fundamental attitudes of, for example, a street kid, Tom, might result in the most informative classification for his self-synthesis when he is in the back-alleyways of Edmonton. A general contempt for others leads him to trust nobody and to act intimidating so that other street people don't mug or abuse him and this is good, and

disrespect for property makes him feel much better about not having any. Tom is organically unified insofar as his attitudes, other psychological states, and actions are not only consistent but mutually reinforcing. Now, the fundamental principles governing the classifications performed by his self-syntheses can begin to show strain if the information source for synthesis, that is, the sorts of things available for him to synthesize, changes.[17] For example, the integration of Tom into an organization of caring individuals will start to strain his attitudes. He initially keeps the attitudes, but it becomes harder for him to make his life go well in the same way he had been. Others are still intimidated by him and this gives him maybe a little power, but it robs him of genuine friendships. On the street, the opportunity for friendship was less frequent, so the loss was a smaller one; but in the organization, he notices that some of the people around him are becoming friends. Moreover, ultimatums are given to him – certain actions that he would have performed on the street, e.g. knocking out the teeth of an insulter, will result in his expulsion from the group (and say this has negative expected utility because as a member of the group he receives certain benefits). With constraints like this, e.g. the threat of discipline, group expectations, etc., actions directly recommended by the street attitude, e.g. outbursts of anger, drug use, gang mentality, are no longer in Tom's rational self-interest. He *can* maintain the street attitude and still act in his self-interest, but this requires something analogous to introducing clauses into what was before a straight-forward plan of action e.g. act intimidating *unless* a supervisor is watching. As more clauses are introduced for more and more situations (hopefully this organization is a broad one), then the street attitudes, although fundamental, might start to do poorly in ordering the priorities and

[17] It is not necessary that his environment change. As I point out above, if a new idea alters a person's mental landscape, this can trigger a self-revolution without any change in the environment.

beliefs of the boy into an organic unity so that his life would be better orchestrated by other fundamental attitudes. At this point, his classification is no longer an informative one because the street attitudes just complicate the mapping of his posterior attitudes, which could be better ordered by new fundamental attitudes so that more of them are better satisfied than before. At this point, his attitudes are vulnerable to revolution but not yet overtaken.

As long as Tom's fundamental attitudes are the street attitudes, according to the mere expected utility of his options it will never be in his rational self-interest to adopt new fundamental attitudes.[18] Therefore, if Tom is to achieve new fundamental attitudes, he must come to select them based upon their symbolic utility. In the optimal scenario, these new attitudes allow him to survive optimally both on the street *and* when integrated in society. The example of Tom demonstrates what I mean by 'response' when claiming that, in order for self-identity to be preserved over self-revolutions, the fundamental cares of the post-revolutionary self-synthesis must be a response to inadequacies in the pre-revolutionary self-syntheses' ability to synthesize the most informative classification. If a person undergoes a revolutionary self-synthesis that is not a response to the previous self-identity, e.g. synthesizing as Napoleon or as Harper or as a non-human object, then self-identity is not preserved.

One might object that there is no reason why synthesizing as Napolean, Harper, or a non-human object could not be a response to inadequacies in the prior self-syntheses. I have two answers to this objection. First, it is very unlikely, at any point of any person's life, that identifying as a person that has lived or is living a life apart from the life one has

[18] Unless the street attitudes are self-defeating, but nothing suggests they have to be.

thus far lived, or as an object that has existed or is existing apart from the existence in which one has thus far partaken, will constitute the most suitable response to the inadequacies of an identity that is attempting to continue itself. This is because these other persons or objects have come to be and have themselves been formed in response to different environmental pressures than one's own history. Their admirable qualities are such relative to different sorts of situations than the ones with which one's prior identity has been having problems e.g. Ghandi's fundamental attitudes probably wouldn't organically unify the resources at the disposal of my current identity's self-synthesis as much as a set of fundamental attitudes uniquely fitted to the situation. Since no two situations are ever exactly alike, identifying as another person will never constitute the best response to the inadequacies of one's prior self-identity given one's current environment. This leads to my second answer, which is that one can never actually identify as another person or as another object (should an object possess the power of self-synthesis) because one can never perfectly recreate the self-syntheses that initially fashioned the identity one is attempting to copy, which occurred in necessarily different contexts than will one's own. These considerations do, however, shed light upon the relative adequacy of role-models. Those people that have performed self-identifications in situations like one's own and with success will *to those extents* be good role-models, but will never provide exceptionless rules for action.

A further possible objection is that, even granted that self-revolutions occur, nothing is gained by viewing them as preservations of personhood. This is Parfit's position that if personal identity is not what matters in situations where it is hard to discern, then this is a defeat of even its practical justification. My response to this,

implicit in the discussion above, is that the gain is one of motivation. If a person were to view even a progressive revolutionary self-synthesis as the death of self-identity, she would not care for it in a self-synthesis upon the Platonic Glasses mode of caring and it would have no symbolic utility. Viewing deep changes of attitude as, necessarily, breaks in self-identity means that certain drastic changes of individuals that are both good for that individual and for society, like prisoner and drug rehabilitation, are types of personal identity suicide. Even if Parfit is right and self-identity is not what matters, that is, even if relation R is all that matters, then since relation R does not hold over revolutionary self-syntheses, Parfit's reductionist perspective eliminates any motivation for the characters involved to *want* to change. On the other hand, if we view these revolutions (the value-making ones) not as changes in personhood but as progresses of personhood, then despite the lack of any physical or psychological continuity, the person involved will be motivated to undergo them with effort.

Still, who is to say that such a strategy does not trick those whom it reforms so that they commit identity-suicide unknowingly under the impression that they were becoming better people? Parfit's interpretation of Nozick's view has this implication. He writes, "Even though it is not true that we are beings whose continued existence must be all-or-nothing, it can be rational to care about our identity as if this was true" (p. 479). He accuses Nozick of 'wishful thinking' but tolerates it given the distinction between practical and theoretical reason. He claims that the Platonic Glasses mode of caring is theoretically irrational, but can be practically rational if it results in enough satisfaction.

I will argue with Parfit that it is not always rational to care about one's identity as if it is all-or-nothing. The view that it is always rational to do so is compatible with

Nozick's position, but so is the view that it is only sometimes rational to do so. I am arguing that it is only sometimes rational to do so and that, contrary to Parfit, it can be theoretically rational to do so. It becomes theoretically rational to do so to the extent that one begins, and continues to live, a valuable life.

IX. Rationality and Reductionism

On Nozick's account of rationality, self-identifying on the Platonic glasses mode of caring can be theoretically rational even if reductionism is true. Nozick defends an instrumentalist conception of rationality but holds that there are other legitimate modes of rationality including the evidential and the symbolic (1993, p. 139). Since Nozick is not a reductionist about personal identity but holds that it is established through self-synthesis, and since identity in a self-identification is both a statement of identity and a predication, one can identify as a determinate entity and at the same time *make true* the belief that one is a determinate identity. Thus, the statement that one is a determinate identity can be true while the incompatible alternative is false because, on Nozick's account, its proposal can make itself true.

Parfit's reductionism motivates his charge that the Platonic Glasses mode of caring is theoretically irrational. Since, on this view, a person is just particles subsisting in more or less of the same structure over time, for this structure to identify as an entity beyond the flux of the very particles that constitute it is for it to err. On this construal, self-identification is a bottom-up phenomenon. It can only occur in the direction, 'biological to psychological.'

However, reductionism does not have to constrain a person's beliefs about herself in this way if these beliefs can be made true through self-identification. Although neurology and biology determine and constrain a person's thoughts at least to a certain extent, what one thinks constrains and determines one's neurology and biology. Consider, for example, a person who thinks he is embarrassed so that his cheeks turn red, he perspires, and his hands tremble. This instance does not depend upon downward causation in any quasi-mystical sense, but perhaps in terms of certain feedback mechanisms. Presumably, when a person starts thinking differently, this thought manifests or corresponds to, (pick your theory), analogous changes, however minor, in the structure of her constituent parts. The question is whether changes in the meta-processes (self-syntheses) regulating a person's self-identity can arise from out of normal processes within the self-identity, which would, in effect, result in a different set of normal processes. For Parfit, this cannot occur; either one's thoughts (a subclass of one's neurological processes) conform to the regularities of the entire class of one's neurological processes or they are wrong; however, on Nozick's construal, it is a type of malfunctioning if the regularities of one's entire class of neurological processes and those of the subclass of these that constitute one's psychological processes fail to engage in the right sort of interactive relationship. A person's neurologically realized psychological thoughts can change for her the *sorts* of thoughts she *can* neurologically realize or the thoughts that he tends to realize. Perhaps the psychological information changes the thoughts a person can neurologically realize for the worse, e.g. after thoughts about gambling, a person can't stop thinking thoughts that result in his giving money away—this is an example of the *wrong* sort of interactive relationship. The right sort of

interactive relationship, according to Nozick, is one that tracks value. This occurs if, for example, a person thinks a thought that results in a tendency or disposition to think more organically unified thoughts than before or think thoughts that lead one to live a more organically unified life than before.

Nozick points out that if ethical behaviour increases inclusive fitness through the very aspects that make it ethical, then it is conceivable how mental processes that 'track value' have been selected for (1981 p. 346). He claims that, just as advanced mathematical knowledge arose from a skill initially selected for relatively minor utilities, "the capacity to recognize ethical truths unveils surprising structures, convolutions, refinements, modulations and asymmetries" (pp. 346-347). Reflective individuals, his story goes, pondering these features, then gave them weights unassociated with selective pressures (p. 347).

Given this notion of value, there is a sense in which a person's beliefs about himself, the truth or falsity of which is determined by that person's self-syntheses, can be held accountable to theoretical rationality. Parfit's charge against Nozick of theoretical irrationality will stick if it is always necessary to be theoretically rational in certain aspects of life. If one's value is one's degree of organic unity, as I have argued, then the more organically unified one is the more one ought to want to preserve oneself and build further upon oneself as a determinate entity. Furthermore, it is irrational to engage in practical pursuits that denigrate one's organic unity because this is actually a negation of personhood and it would be, quite literally, a self-defeating task to negate oneself. Moreover, if one cares about oneself as less determinate than one ought to, given one's

organic unity, then this is irrational; if one cares about oneself as more determinate than one ought to, given one's organic unity, then this is irrational.

If one loses organic unity, then, in Parfitian language, one's psychological continuity begins to break down. However, I do not want to say with Parfit that a loss of psychological continuity corresponds to there being less similarity between stages of me; rather, a loss of psychological continuity corresponds to there being less and less of an entity for there to be stages between. There is still a spectrum of continuity but it covers the relevance, not the degree, of continuity. When Parfit claims that relation R is all that matters, he claims that the only thing relevant to questions about continuity is the extent to which psychological connectedness is maintained. However, there can be many psychological connections between two stages of a person but, given the nature of the stages, the connections can be relatively unimportant. Consider a person who can only ever think about butter. He thinks about butter in the morning, in the afternoon, and in the evening: this person *is* continuing but there isn't much *to* this person, the thing that is continuing is less of an entity than a more capacious person would be.

When I am an organic unity and an embodiment of value there is a tangible sense in which it can be in my best interest to continue determinately because there is much to continue. On the other hand, if I am barely a person to begin with, if I merely ebb and flow with my environment like the rivers and trees, then I care little about my continuity because there is not that much to be continued and not that much to do the caring. This is not to say that a person cannot sacrifice his personal identity for practical considerations like, for an outrageous example, becoming addicted to prescription pills at gunpoint to save the life of one's daughter, but only that a certain theoretical rigor distinguishing

between appropriate and inappropriate practical aims must receive constant maintenance for the preservation of one's self-identity. As to live and to physically grow, it is always necessary to maintain certain bodily functions, the flow of oxygenated blood, brain activity, etc., so it is necessary to maintain, I argue, certain rational functions to maintain and increase the determinacy of one's personal-identity.

I have been claiming that personal identity is preserved over progressive revolutionary self-syntheses. One might object that if one's identity only becomes more determinate *after* a progressive revolution, which is only after the worth of preserving one's identity increases, and if it is not determinate before the revolution, then the sense in which there is preservation of self-identity over self-revolutions is ambiguous.

Once one's identity is determinate, the determinacy his past identities were interpreted to have before any knowledge of their future is no longer relevant. The entity constituted by a person's self-identity and all that it symbolizes with its metric entails, given the objects available for synthesis, its closest continuers as well as its closest predecessors as parts of itself. This features in Nozick's closest continuer theory when he adds the condition that a stage *Y* that is the closest continuer of stage *X* can only be the same object as *X* if *X* is *Y*'s closest predecessor (p. 42). This closest predecessor condition was meant to guarantee that the closest continuer 'stem from' that which it is most similar to (p. 41).

While explaining his notion of the informative classification, Nozick notes that the distinction between transverse and longitudinal identity is largely artificial and compares them to the terms, 'timelike' and 'spacelike' in relativity theory (pp. 85-86). One reason the distinction is artificial is because "if causal interconnections among parts

are part of closeness in the transverse metric space then, via causal connection, time already enters into transverse identity" (p. 85). Transverse identity here implies a span of longitudinal identity because causal connections occur in time. Likewise, symbolic connections can occur in time and most of them do. The degree of organic unity constituting the extent to which a person is determinate is in part constituted by symbolic connections between her various stages whether or not they are psychologically connected. In actuality, and once the later identity is known to be determinate, the prior identity can be rightfully interpreted as a *part* of the later determinate entity because it contributes to its significance. For example, a recovered drug addict's attitude toward life gains meaning from the fact that she was once a drug addict, even if the attitudes of the addict were once fundamental. The fact that she once had the fundamental attitudes of a drug addict will, for instance, affect the symbolic utilities she assigns to things. For example, after recovery, she might refuse a sociable drink at a dinner party, even though the consequences of accepting it would have maximized her expected utility, because strict sobriety has taken on symbolic meaning for her. The earlier entity is a symbolically connected part of the later entity and it contributes to the overall organic unity and value of that later entity.

IX. Conclusion

An analysis of Nozick's closest continuer theory of personal identity has shown that, without some further assumption, the closest continuer theory of personal identity cannot claim that radical changes in persons preserve personhood. The right interpretation on Nozick's objective theory of value fixes this problem by providing the tools to distinguish valuable from invaluable radical changes in personhood. Self-identity

is determinate over changes in personhood that preserve or increase one's value. Under these added value-preserving or value-increasing conditions, the closest continuer theory's commitment to the determinacy of personal identity over radical changes survives Parfit's reductionist objection that personal identity is only a matter of degree. I have shown that personal identity is not equally a matter of degree in all cases, but that it can become a more determinate entity, the continuation of which is less a matter of degree.

BIBLIOGRAPHY

Bean, Henry, dir. *The Believer*. Perfs. Ryan Gosling. Fuller Films, 2001.

The Compact Oxford English Dictionary. 3rd edition. www.askoxford.com, 2005.

Galatians, *The Holy Bible*. The New International Version, 1984.

Little, Malcolm X. *The Autobiography of Malcolm X*. Grove Press Inc.: New York, 1965.

Nozick, Robert. *Philosophical Explanations*. Cambridge, MA: Harvard University Press, 1981.

Nozick, Robert. *The Examined Life*. Simon & Schuster: New York, 1990.

Nozick, Robert. *The Nature of Rationality*. Princeton University Press: Princeton, 1993.

Parfit, Derek. *Climbing the Mountain*. Forthcoming.

Parfit, Derek. *Reasons and Persons*. Oxford: Clarendon Press, 1984.

Putnam, Hilary. "The Analytic and the Synthetic." In *Minnesota Studies in the Philosophy of Science*, vol. 3, ed. H. Feigl and G. Maxwell. Minneapolis: University of Minnesota Press, 1962.

Wimsatt, W. C. *Re-Engineering Philosophy for Limited Beings*. Harvard University Press. Cambridge, 2007.